

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:06 ; Search time 35.9589 Seconds  
(without alignments)  
55.003 Million cell updates/sec

Title: US-09-973-473A-21  
Perfect score: 15  
Sequence: 1 YXXXLX 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	68.8	9	6	ABP83755 HLA prote
2	11	68.8	9	6	ABP84060 HLA prote
3	11	68.8	9	6	ABP84934 HLA prote
4	11	68.8	9	6	ABP84936 HLA prote
5	11	68.8	9	6	ABP85536 HLA prote
6	11	68.8	9	6	ABP85937 HLA prote
7	11	68.8	9	6	ABP86587 HLA prote
8	11	68.8	9	6	ABP87076 HLA prote
9	11	68.8	9	6	ABP87077 HLA prote
10	11	68.8	9	6	ABP88907 HLA prote
11	11	68.8	9	6	ABP89307 HLA prote
12	11	68.8	9	6	ABP84056 HLA prote
13	11	68.8	9	6	ABP84643 HLA prote
14	11	68.8	9	6	ABP85166 HLA prote
15	11	68.8	9	6	ABP86592 HLA prote
16	11	68.8	9	6	ABP87280 HLA prote
17	11	68.8	9	6	ABP88388 HLA prote
18	11	68.8	9	6	ABP88900 HLA prote
19	11	68.8	9	6	ABP88905 HLA prote
20	11	68.8	9	6	ABP89431 HLA prote
21	11	68.8	9	6	ABP83762 HLA prote
22	11	68.8	9	6	ABP84937 HLA prote
23	11	68.8	9	6	ABP85237 HLA prote
24	11	68.8	9	6	ABP87566 HLA prote
25	11	68.8	9	6	ABP87601 HLA prote

26	11	68.8	9	6	ABP88821	Abp88821 HLA prote
27	11	68.8	9	6	ABP88901	Abp88901 HLA prote
28	11	68.8	9	6	ABP89396	Abp89396 HLA prote
29	11	68.8	9	6	ABP90314	Abp90314 HLA prote
30	11	68.8	9	6	ABP90317	Abp90317 HLA prote
31	11	68.8	9	6	ABP84640	Abp84640 HLA prote
32	11	68.8	9	6	ABP84831	Abp84831 HLA prote
33	11	68.8	9	6	ABP85528	Abp85528 HLA prote
34	11	68.8	9	6	ABP86101	Abp86101 HLA prote
35	11	68.8	9	6	ABP88448	Abp88448 HLA prote
36	11	68.8	9	6	ABP88453	Abp88453 HLA prote
37	11	68.8	9	6	ABP88526	Abp88526 HLA prote
38	11	68.8	9	6	ABP88937	Abp88937 HLA prote
39	11	68.8	9	6	ABP89003	Abp89003 HLA prote
40	11	68.8	9	6	ABP89463	Abp89463 HLA prote
41	11	68.8	9	6	ABP90547	Abp90547 HLA prote
42	11	68.8	9	6	ABP84042	Abp84042 HLA prote
43	11	68.8	9	6	ABP84638	Abp84638 HLA prote
44	11	68.8	9	6	ABP85239	Abp85239 HLA prote
45	11	68.8	9	6	ABP85768	Abp85768 HLA prote

ALIGNMENTS

RESULT 1  
ABP83755  
ID ABP83755 standard; peptide; 9 AA.

XX ABP83755;  
AC  
XX 28-MAR-2003 (first entry)  
DT  
XX  
XX HLA protein 121P2A3 peptide #110.  
DE  
XX  
XX Human; 121P2A3; cytosolic; immunostimulant; vaccine; SSH;  
KW humoral immune response; cellular immune response;  
KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200283068-A2.  
PN  
XX  
XX 24-OCT-2002.  
PD  
XX  
XX 09-APR-2002; 2002WO-US011359.  
PF  
XX  
XX 10-APR-2001; 2001US-0282739P.  
PR  
XX  
XX 25-APR-2001; 2001US-0286630P.  
PR  
XX  
XX 22-JUN-2001; 2001US-0300373P.  
PA  
XX  
XX (AGEN-) AGENSYS INC.  
PI  
XX  
XX Challita-Eid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
FI Afar DEH, Saffran D, Morrison K, Morrison KK, Ge W, Jakobovits A;  
DR  
XX  
XX WPI; 2003-092956/08.  
XX  
XX New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
XX versus cancerous tissues.  
XX  
XX Claim 13; Page 130; 362pp; English.

XX  
XX The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to

CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure  
CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
XX represent peptides from the 121P2A3 variants of the invention  
XX Sequence 9 AA;

Query Match 68.8%; Score 11; DB 6; Length 9;  
Best Local Similarity 33.3%; Pred. No. 1.3e+06; Indels 0; Gaps 0;  
Matches 2; Conservative 0; Mismatches 4;

QY 1 YXXXXL 6  
Db 1 YSTTTL 6

RESULT 2  
ABP84060  
ID ABP84060 standard; peptide; 9 AA.  
XX  
AC ABP84060;  
XX  
DT 28-MAR-2003 (first entry)  
XX  
DE HLA protein 121P2A3 peptide #415.  
XX  
XX Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;  
KW humoral immune response; cellular immune response;  
KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
XX  
OS Homo sapiens.  
XX  
PN WO200283068-A2.  
XX  
PD 24-OCT-2002.  
XX  
PF 09-APR-2002; 2002WO-US011359.  
XX  
PR 10-APR-2001; 2001US-0282739P.  
PR 25-APR-2001; 2001US-0286630P.  
PR 22-JUN-2001; 2001US-0300373P.  
XX  
PA (AGEN-) AGENSYS INC.  
XX  
PI Challita-Eid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
XX  
DR WPI; 2003-092956/08.  
XX  
XX New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
PT versus cancerous tissues.  
XX  
PS Claim 13; Page 136; 362pp; English.  
XX  
XX The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to  
CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure

CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
XX represent peptides from the 121P2A3 variants of the invention  
XX Sequence 9 AA;

Query Match 68.8%; Score 11; DB 6; Length 9;  
Best Local Similarity 33.3%; Pred. No. 1.3e+06; Indels 0; Gaps 0;  
Matches 2; Conservative 0; Mismatches 4;

QY 1 YXXXXL 6  
Db 2 YSTTTL 7

RESULT 3  
ABP84934  
ID ABP84934 standard; peptide; 9 AA.  
XX  
AC ABP84934;  
XX  
DT 28-MAR-2003 (first entry)  
XX  
DE HLA protein 121P2A3 peptide #1289.  
XX  
XX Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;  
KW humoral immune response; cellular immune response;  
KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
XX  
OS Homo sapiens.  
XX  
PN WO200283068-A2.  
XX  
PD 24-OCT-2002.  
XX  
PF 09-APR-2002; 2002WO-US011359.  
XX  
PR 10-APR-2001; 2001US-0282739P.  
PR 25-APR-2001; 2001US-0286630P.  
PR 22-JUN-2001; 2001US-0300373P.  
XX  
PA (AGEN-) AGENSYS INC.  
XX  
PI Challita-Eid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
XX  
DR WPI; 2003-092956/08.  
XX  
XX New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
PT versus cancerous tissues.  
XX  
PS Claim 13; Page 153; 362pp; English.  
XX  
XX The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to  
CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure  
CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP83646 - ABP95595

CC represent peptides from the 121P2A3 variants of the invention  
XX  
SQ Sequence 9 AA;

Query Match 68.8%; Score 11; DB 6; Length 9;  
Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6  
|  
Db 4 YSTTTL 9

RESULT 4  
ABP84936  
ID ABP84936 standard; peptide; 9 AA.

XX AC ABP84936;

XX DT 28-MAR-2003 (first entry)

XX DE HLA protein 121P2A3 peptide #1291.

XX KW Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;

XX KW humoral immune response; cellular immune response;

XX KW suppression subtractive hybridisation; HLA; human leukocyte antigen.

XX OS Homo sapiens.

XX PN WO200283068-A2.

XX PD 24-OCT-2002.

XX PF 09-APR-2002; 2002WO-US011359.

XX PR 10-APR-2001; 2001US-0282739P.

XX PR 25-APR-2001; 2001US-0286630P.

XX PR 22-JUN-2001; 2001US-0300373P.

XX PA (AGEN-) AGENSYS INC.

XX PI Challita-Bid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;

XX PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;

XX DR WPI; 2003-092956/08.

XX PT New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
PT versus cancerous tissues.

XX PS Claim 13; Page 153; 362pp; English.

XX CC The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to  
CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure  
CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
CC represent peptides from the 121P2A3 variants of the invention

XX SQ Sequence 9 AA;

Query Match  
Best Local Similarity 33.3%;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Query Match 68.8%; Score 11; DB 6; Length 9;  
Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6  
|  
Db 2 YSTTTL 7

RESULT 5

ABP85536  
ID ABP85536 standard; peptide; 9 AA.

XX AC ABP85536;

XX DT 28-MAR-2003 (first entry)

XX DE HLA protein 121P2A3 peptide #1891.

XX KW Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;

XX KW humoral immune response; cellular immune response;

XX KW suppression subtractive hybridisation; HLA; human leukocyte antigen.

XX OS Homo sapiens.

XX PN WO200283068-A2.

XX PD 24-OCT-2002.

XX PF 09-APR-2002; 2002WO-US011359.

XX PR 10-APR-2001; 2001US-0282739P.

XX PR 25-APR-2001; 2001US-0286630P.

XX PR 22-JUN-2001; 2001US-0300373P.

XX PA (AGEN-) AGENSYS INC.

XX PI Challita-Bid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;

XX PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;

XX DR WPI; 2003-092956/08.

XX PT New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
PT versus cancerous tissues.

XX PS Claim 13; Page 165; 362pp; English.

XX CC The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to  
CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure  
CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
CC represent peptides from the 121P2A3 variants of the invention

XX SQ Sequence 9 AA;

Query Match  
Best Local Similarity 33.3%;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
 Db 2 YSTTTL 7

RESULT 6  
 ABP85937  
 ID ABP85937 standard; peptide; 9 AA.  
 AC  
 XX ABP85937;  
 XX  
 XX 28-MAR-2003 (first entry)  
 DT  
 XX  
 XX HLA protein 121P2A3 peptide #2292.  
 DE  
 XX Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;  
 KW humoral immune response; cellular immune response;  
 KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200283068-A2.  
 PN  
 XX 24-OCT-2002.  
 PD  
 XX 09-APR-2002; 2002WO-US011359.  
 PF  
 XX 10-APR-2001; 2001US-0282739P.  
 PR  
 XX 25-APR-2001; 2001US-0286630P.  
 PR  
 XX 22-JUN-2001; 2001US-0300373P.  
 XX  
 XX (AGEN-) AGENSYS INC.  
 PA  
 XX Challita-Eid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
 PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
 PI WPI; 2003-092956/08.  
 DR  
 XX New composition comprising a substance that modulates the status of  
 PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
 PT responses or in assessing the status of 121P2A3 gene products in normal  
 PT versus cancerous tissues.  
 XX  
 XX Claim 13; Page 175; 362pp; English.  
 PS  
 XX The invention relates to a novel composition comprising a substance that  
 CC modulates the status of a protein, 121P2A3. The composition of the  
 CC invention has cytostatic and immunostimulant activity, and is useful as a  
 CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
 CC eliciting humoral or cellular immune response. The polynucleotides are  
 CC useful for characterising cytogenetic abnormalities of this chromosomal  
 CC locus, as tools that can be used to delineate cytogenetic abnormalities  
 CC in the chromosomal region that encodes 121P2A3 that may contribute to  
 CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
 CC in normal versus cancerous tissues. The proteins are useful for  
 CC generating and characterising domain-specific antibodies, for identifying  
 CC agents or cellular factors that bind to 121P2A3 or a particular structure  
 CC domain, and in various therapeutic and diagnostic contexts, including  
 CC cancer vaccines. The antibodies or T cells reactive with the product are  
 CC useful in passive or active immunisation, and in imaging methodologies  
 CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
 CC represent peptides from the 121P2A3 variants of the invention  
 XX  
 SQ Sequence 9 AA;

Query Match 68.8%; Score 11; DB 6; Length 9;  
 Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
 Db 3 YSTTTL 8

RESULT 7  
 ABP86587  
 ID ABP86587 standard; peptide; 9 AA.  
 AC  
 XX ABP86587;  
 XX  
 XX 28-MAR-2003 (first entry)  
 DT  
 XX  
 XX HLA protein 121P2A3 peptide #2942.  
 DE  
 XX Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;  
 KW humoral immune response; cellular immune response;  
 KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200283068-A2.  
 PN  
 XX 24-OCT-2002.  
 PD  
 XX 09-APR-2002; 2002WO-US011359.  
 PF  
 XX 10-APR-2001; 2001US-0282739P.  
 PR  
 XX 25-APR-2001; 2001US-0286630P.  
 PR  
 XX 22-JUN-2001; 2001US-0300373P.  
 XX  
 XX (AGEN-) AGENSYS INC.  
 PA  
 XX Challita-Bid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
 PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
 PI WPI; 2003-092956/08.  
 DR  
 XX New composition comprising a substance that modulates the status of  
 PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
 PT responses or in assessing the status of 121P2A3 gene products in normal  
 PT versus cancerous tissues.  
 XX  
 XX Claim 13; Page 182; 362pp; English.  
 PS  
 XX The invention relates to a novel composition comprising a substance that  
 CC modulates the status of a protein, 121P2A3. The composition of the  
 CC invention has cytostatic and immunostimulant activity, and is useful as a  
 CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
 CC eliciting humoral or cellular immune response. The polynucleotides are  
 CC useful for characterising cytogenetic abnormalities of this chromosomal  
 CC locus, as tools that can be used to delineate cytogenetic abnormalities  
 CC in the chromosomal region that encodes 121P2A3 that may contribute to  
 CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
 CC in normal versus cancerous tissues. The proteins are useful for  
 CC generating and characterising domain-specific antibodies, for identifying  
 CC agents or cellular factors that bind to 121P2A3 or a particular structure  
 CC domain, and in various therapeutic and diagnostic contexts, including  
 CC cancer vaccines. The antibodies or T cells reactive with the product are  
 CC useful in passive or active immunisation, and in imaging methodologies  
 CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
 CC represent peptides from the 121P2A3 variants of the invention  
 XX  
 SQ Sequence 9 AA;

Query Match 68.8%; Score 11; DB 6; Length 9;  
 Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
 Db 3 YSTTTL 8

RESULT 8  
 ABP87076  
 ID ABP87076 standard; peptide; 9 AA.



```
XX AC ABP87076;
XX DT 28-MAR-2003 (first entry)
XX DE HLA protein 121P2A3 peptide #3431.
XX KW Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;
XX KW humoral immune response; cellular immune response;
XX KW suppression subtractive hybridisation; HLA; human leukocyte antigen.
XX OS Homo sapiens.
XX PN WO200283068-A2.
XX PD 24-OCT-2002.
XX PF 09-APR-2002; 2002WO-US011359.
XX PR 10-APR-2001; 2001US-0282739P.
XX PR 25-APR-2001; 2001US-0286630P.
XX PR 22-JUN-2001; 2001US-0300373P.
XX PA (AGEN-) AGENSYS INC.
XX PI Challita-Bid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;
XX PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;
XX DR WPI; 2003-092956/08.
XX PS New composition comprising a substance that modulates the status of
XX PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune
XX PT responses or in assessing the status of 121P2A3 gene products in normal
XX PT versus cancerous tissues.
XX PS Claim 13; Page 186; 362pp; English.
XX CC The invention relates to a novel composition comprising a substance that
XX CC modulates the status of a protein, 121P2A3. The composition of the
XX CC invention has cytostatic and immunostimulant activity, and is useful as a
XX CC vaccine. The 121P2A3 proteins and polynucleotides are useful for
XX CC eliciting humoral or cellular immune response. The polynucleotides are
XX CC useful for characterising cytogenetic abnormalities of this chromosomal
XX CC locus, as tools that can be used to delineate cytogenetic abnormalities
XX CC in the chromosomal region that encodes 121P2A3 that may contribute to
XX CC malignant phenotype, and in assessing the status of 121P2A3 gene products
XX CC in normal versus cancerous tissues. The proteins are useful for
XX CC generating and characterising domain-specific antibodies, for identifying
XX CC agents or cellular factors that bind to 121P2A3 or a particular structure
XX CC domain, and in various therapeutic and diagnostic contexts, including
XX CC cancer vaccines. The antibodies or T cells reactive with the product are
XX CC useful in passive or active immunisation, and in imaging methodologies
XX CC for the management of cancer. The sequences shown in ABP83646 - ABP95595
XX CC represent peptides from the 121P2A3 variants of the invention
XX SQ Sequence 9 AA;
Query Match 68.8%; Score 11; DB 6; Length 9;
Best Local Similarity 33.3%; Pred. No. 1.3e+06;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 YXXXXL 6
DB 3 YSTTTL 8
RESULT 9
ABP87077
ID ABP87077 standard; peptide; 9 AA.
XX AC ABP87077;
XX DT 28-MAR-2003 (first entry)
```

```
XX DE HLA protein 121P2A3 peptide #3432.
XX KW Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;
XX KW humoral immune response; cellular immune response;
XX KW suppression subtractive hybridisation; HLA; human leukocyte antigen.
XX OS Homo sapiens.
XX PN WO200283068-A2.
XX PD 24-OCT-2002.
XX PF 09-APR-2002; 2002WO-US011359.
XX PR 10-APR-2001; 2001US-0282739P.
XX PR 25-APR-2001; 2001US-0286630P.
XX PR 22-JUN-2001; 2001US-0300373P.
XX PA (AGEN-) AGENSYS INC.
XX PI Challita-Bid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;
XX PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;
XX DR WPI; 2003-092956/08.
XX PS New composition comprising a substance that modulates the status of
XX PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune
XX PT responses or in assessing the status of 121P2A3 gene products in normal
XX PT versus cancerous tissues.
XX PS Claim 13; Page 186; 362pp; English.
XX CC The invention relates to a novel composition comprising a substance that
XX CC modulates the status of a protein, 121P2A3. The composition of the
XX CC invention has cytostatic and immunostimulant activity, and is useful as a
XX CC vaccine. The 121P2A3 proteins and polynucleotides are useful for
XX CC eliciting humoral or cellular immune response. The polynucleotides are
XX CC useful for characterising cytogenetic abnormalities of this chromosomal
XX CC locus, as tools that can be used to delineate cytogenetic abnormalities
XX CC in the chromosomal region that encodes 121P2A3 that may contribute to
XX CC malignant phenotype, and in assessing the status of 121P2A3 gene products
XX CC in normal versus cancerous tissues. The proteins are useful for
XX CC generating and characterising domain-specific antibodies, for identifying
XX CC agents or cellular factors that bind to 121P2A3 or a particular structure
XX CC domain, and in various therapeutic and diagnostic contexts, including
XX CC cancer vaccines. The antibodies or T cells reactive with the product are
XX CC useful in passive or active immunisation, and in imaging methodologies
XX CC for the management of cancer. The sequences shown in ABP83646 - ABP95595
XX CC represent peptides from the 121P2A3 variants of the invention
XX SQ Sequence 9 AA;
Query Match 68.8%; Score 11; DB 6; Length 9;
Best Local Similarity 33.3%; Pred. No. 1.3e+06;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 YXXXXL 6
DB 2 YSTTTL 7
RESULT 10
ABP88907
ID ABP88907 standard; peptide; 9 AA.
XX AC ABP88907;
XX DT 28-MAR-2003 (first entry)
XX DE HLA protein 121P2A3 peptide #5262.
XX KW Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;
```

KW humoral immune response; cellular immune response;  
KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
OS  
OS Homo sapiens.  
XX WO200283068-A2.  
PN  
XX  
XX 24-OCT-2002.  
XX  
XX 09-APR-2002; 2002WO-US011359.  
XX  
XX 10-APR-2001; 2001US-0282739P.  
PR 25-APR-2001; 2001US-0286630P.  
PR 22-JUN-2001; 2001US-0300373P.  
XX  
XX (AGEN-) AGENSYS INC.  
PA  
XX Challita-Eid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
XX WPI; 2003-092956/08.  
DR  
XX  
XX New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
PT versus cancerous tissues.  
XX  
XX Claim 13; Page 204; 362pp; English.  
PS  
XX The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to  
CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure  
CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
CC represent peptides from the 121P2A3 variants of the invention  
XX Sequence 9 AA;  
SQ  
Query Match 68.8%; Score 11; DB 6; Length 9;  
Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 YXXXXL 6  
Db 2 YSTTL 7  
RESULT 11  
ABP89307  
ID ABP89307 standard; peptide; 9 AA.  
XX  
AC ABP89307;  
XX  
XX 28-MAR-2003 (first entry)  
DT  
XX HLA protein 121P2A3 peptide #5662.  
DE  
XX Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;  
KW humoral immune response; cellular immune response;  
KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
XX  
XX Homo sapiens.

XX WO200283068-A2.  
PN  
XX  
XX 24-OCT-2002.  
XX  
XX 09-APR-2002; 2002WO-US011359.  
XX  
XX 10-APR-2001; 2001US-0282739P.  
PR 25-APR-2001; 2001US-0286630P.  
PR 22-JUN-2001; 2001US-0300373P.  
XX  
XX (AGEN-) AGENSYS INC.  
PA  
XX Challita-Eid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
XX WPI; 2003-092956/08.  
DR  
XX  
XX New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
PT versus cancerous tissues.  
XX  
XX Claim 13; Page 208; 362pp; English.  
PS  
XX The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to  
CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure  
CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
CC represent peptides from the 121P2A3 variants of the invention  
XX Sequence 9 AA;  
SQ  
Query Match 68.8%; Score 11; DB 6; Length 9;  
Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 YXXXXL 6  
Db 1 YSTTL 6  
RESULT 12  
ABP84056  
ID ABP84056 standard; peptide; 9 AA.  
XX  
AC ABP84056;  
XX  
XX 28-MAR-2003 (first entry)  
DT  
XX HLA protein 121P2A3 peptide #411.  
DE  
XX Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;  
KW humoral immune response; cellular immune response;  
KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
XX  
XX Homo sapiens.  
XX WO200283068-A2.  
PN  
XX 24-OCT-2002.

XX PF 09-APR-2002; 2002WO-US011359.  
 XX XX  
 PR 10-APR-2001; 2001US-0282739P.  
 PR 25-APR-2001; 2001US-0286630P.  
 PR 22-JUN-2001; 2001US-0300373P.  
 XX XX  
 PA (AGEN-) AGENSYS INC.  
 XX  
 PI Challita-Bid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
 PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
 XX  
 DR WPI; 2003-092956/08.  
 XX  
 PT New composition comprising a substance that modulates the status of  
 PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
 PT responses or in assessing the status of 121P2A3 gene products in normal  
 PT versus cancerous tissues.  
 XX  
 PS Claim 13; Page 136; 362pp; English.  
 CC  
 CC The invention relates to a novel composition comprising a substance that  
 CC modulates the status of a protein, 121P2A3. The composition of the  
 CC invention has cytostatic and immunostimulant activity, and is useful as a  
 CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
 CC eliciting humoral or cellular immune response. The polynucleotides are  
 CC useful for characterising cytogenetic abnormalities of this chromosomal  
 CC locus, as tools that can be used to delineate cytogenetic abnormalities  
 CC in the chromosomal region that encodes 121P2A3 that may contribute to  
 CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
 CC in normal versus cancerous tissues. The proteins are useful for  
 CC generating and characterising domain-specific antibodies, for identifying  
 CC agents or cellular factors that bind to 121P2A3 or a particular structure  
 CC domain, and in various therapeutic and diagnostic contexts, including  
 CC cancer vaccines. The antibodies or T cells reactive with the product are  
 CC useful in passive or active immunisation, and in imaging methodologies  
 CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
 CC represent peptides from the 121P2A3 variants of the invention  
 XX  
 SQ Sequence 9 AA;  
 Query Match 68.8%; Score 11; DB 6; Length 9;  
 Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 YXXXXL 6  
 Db 3 YSTTTL 8  
 RESULT 13  
 ABP84643  
 ID ABP84643 standard; peptide; 9 AA.  
 XX  
 AC ABP84643;  
 XX  
 DT 28-MAR-2003 (first entry)  
 XX  
 DE HLA protein 121P2A3 peptide #998.  
 XX  
 KW Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;  
 KW humoral immune response; cellular immune response;  
 KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200283068-A2.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 09-APR-2002; 2002WO-US011359.  
 XX  
 PR 10-APR-2001; 2001US-0282739P.  
 XX

PR 25-APR-2001; 2001US-0286630P.  
 PR 22-JUN-2001; 2001US-0300373P.  
 XX  
 PA (AGEN-) AGENSYS INC.  
 XX  
 PI Challita-Bid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
 PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
 XX  
 DR WPI; 2003-092956/08.  
 XX  
 PT New composition comprising a substance that modulates the status of  
 PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
 PT responses or in assessing the status of 121P2A3 gene products in normal  
 PT versus cancerous tissues.  
 XX  
 PS Claim 13; Page 147; 362pp; English.  
 CC  
 CC The invention relates to a novel composition comprising a substance that  
 CC modulates the status of a protein, 121P2A3. The composition of the  
 CC invention has cytostatic and immunostimulant activity, and is useful as a  
 CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
 CC eliciting humoral or cellular immune response. The polynucleotides are  
 CC useful for characterising cytogenetic abnormalities of this chromosomal  
 CC locus, as tools that can be used to delineate cytogenetic abnormalities  
 CC in the chromosomal region that encodes 121P2A3 that may contribute to  
 CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
 CC in normal versus cancerous tissues. The proteins are useful for  
 CC generating and characterising domain-specific antibodies, for identifying  
 CC agents or cellular factors that bind to 121P2A3 or a particular structure  
 CC domain, and in various therapeutic and diagnostic contexts, including  
 CC cancer vaccines. The antibodies or T cells reactive with the product are  
 CC useful in passive or active immunisation, and in imaging methodologies  
 CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
 CC represent peptides from the 121P2A3 variants of the invention  
 XX  
 SQ Sequence 9 AA;  
 Query Match 68.8%; Score 11; DB 6; Length 9;  
 Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 YXXXXL 6  
 Db 3 YSTTTL 8  
 RESULT 14  
 ABP85166  
 ID ABP85166 standard; peptide; 9 AA.  
 XX  
 AC ABP85166;  
 XX  
 DT 28-MAR-2003 (first entry)  
 XX  
 DE HLA protein 121P2A3 peptide #1521.  
 XX  
 KW Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;  
 KW humoral immune response; cellular immune response;  
 KW suppression subtractive hybridisation; HLA; human leukocyte antigen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200283068-A2.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 09-APR-2002; 2002WO-US011359.  
 XX  
 PR 10-APR-2001; 2001US-0282739P.  
 PR 25-APR-2001; 2001US-0286630P.  
 PR 22-JUN-2001; 2001US-0300373P.  
 XX  
 PA (AGEN-) AGENSYS INC.

XX  
PI Challita-Eid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;  
PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;  
XX  
XX WPI; 2003-092956/08.  
XX  
PT New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
PT versus cancerous tissues.  
XX  
PS Claim 13; Page 158; 362pp; English.  
XX  
CC The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to  
CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure  
CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP95595  
CC represent peptides from the 121P2A3 variants of the invention  
XX  
SQ Sequence 9 AA;

Query Match 68.8%; Score 11; DB 6; Length 9;  
Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| | |  
Db 3 YSTTAL 8

RESULT 15  
ABP86592  
ID ABP86592 standard; peptide; 9 AA.

AC ABP86592;

DT 28-MAR-2003 (first entry)

XX HLA protein 121P2A3 peptide #2947.

DE Human; 121P2A3; cytostatic; immunostimulant; vaccine; SSH;

XX humoral immune response; cellular immune response;

KW suppression subtractive hybridisation; HLA; human leukocyte antigen.

XX Homo sapiens.

XX WO200283068-A2.

XX 24-OCT-2002.

PF 09-APR-2002; 2002WO-US011359.

XX 10-APR-2001; 2001US-0282739P.

PR 25-APR-2001; 2001US-0286630P.

PR 22-JUN-2001; 2001US-0300373P.

XX (AGEN-) AGENSYS INC.

XX Challita-Eid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;

PI Afar DEH, Saffran D, Morrison K, Morrison RK, Ge W, Jakobovits A;

XX

DR WPI; 2003-092956/08.  
XX  
PT New composition comprising a substance that modulates the status of  
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune  
PT responses or in assessing the status of 121P2A3 gene products in normal  
PT versus cancerous tissues.  
XX  
PS Claim 13; Page 182; 362pp; English.  
XX  
CC The invention relates to a novel composition comprising a substance that  
CC modulates the status of a protein, 121P2A3. The composition of the  
CC invention has cytostatic and immunostimulant activity, and is useful as a  
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for  
CC eliciting humoral or cellular immune response. The polynucleotides are  
CC useful for characterising cytogenetic abnormalities of this chromosomal  
CC locus, as tools that can be used to delineate cytogenetic abnormalities  
CC in the chromosomal region that encodes 121P2A3 that may contribute to  
CC malignant phenotype, and in assessing the status of 121P2A3 gene products  
CC in normal versus cancerous tissues. The proteins are useful for  
CC generating and characterising domain-specific antibodies, for identifying  
CC agents or cellular factors that bind to 121P2A3 or a particular structure  
CC domain, and in various therapeutic and diagnostic contexts, including  
CC cancer vaccines. The antibodies or T cells reactive with the product are  
CC useful in passive or active immunisation, and in imaging methodologies  
CC for the management of cancer. The sequences shown in ABP83646 - ABP95595  
CC represent peptides from the 121P2A3 variants of the invention  
XX  
SQ Sequence 9 AA;

Query Match 68.8%; Score 11; DB 6; Length 9;  
Best Local Similarity 33.3%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| | |  
Db 2 YSTTTL 7

Search completed: October 5, 2004, 16:06:28  
Job time : 36.9589 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:00:17 ; Search time 9.4915 Seconds  
(without alignments)  
38.068 Million cell updates/sec

Title: US-09-973-473A-21

Perfect score: 16

Sequence: 1 YXXXXLX 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	11	68.8	11	2	US-08-540-412-123	Sequence 123, App
2	11	68.8	11	2	US-08-540-412-191	Sequence 191, App
3	11	68.8	11	3	US-09-051-342-123	Sequence 123, App
4	11	68.8	11	3	US-09-051-342-191	Sequence 191, App
5	11	68.8	11	3	US-08-468-161-123	Sequence 123, App
6	11	68.8	11	3	US-09-051-759-123	Sequence 123, App
7	11	68.8	11	3	US-09-051-759-191	Sequence 191, App
8	11	68.8	11	5	PCT-US95-08156-123	Sequence 123, App
9	11	68.8	18	6	S219837-5	Patent No. S219837
10	11	68.8	37	3	US-08-905-223-22	Sequence 22, Appl
11	11	68.8	37	4	US-09-247-155-22	Sequence 22, Appl
12	11	68.8	37	4	US-09-663-600A-22	Sequence 22, Appl
13	11	68.8	40	3	US-08-812-586-40	Sequence 40, Appl
14	11	68.8	40	4	US-09-535-832A-37	Sequence 37, Appl
15	11	68.8	64	4	US-09-107-532A-4990	Sequence 4990, Ap
16	11	68.8	76	3	US-08-851-362D-21	Sequence 21, Appl
17	11	68.8	86	4	US-09-252-991A-29383	Sequence 29383, A
18	11	68.8	92	2	US-08-273-146-45	Sequence 45, Appl
19	11	68.8	92	2	US-08-273-146-53	Sequence 53, Appl
20	11	68.8	95	4	US-09-472-087-94	Sequence 94, Appl
21	11	68.8	98	1	US-08-401-908-2	Sequence 2, Appl
22	11	68.8	100	4	US-09-899-896-8	Sequence 8, Appl
23	11	68.8	103	2	US-08-585-585A-5	Sequence 5, Appl
24	11	68.8	103	2	US-08-249-037C-5	Sequence 5, Appl
25	11	68.8	103	2	US-08-788-622B-5	Sequence 5, Appl
26	11	68.8	103	3	US-08-788-621B-5	Sequence 5, Appl
27	11	68.8	104	1	US-07-789-344A-8	Sequence 8, Appl

28	11	68.8	104	1	US-08-276-852-106	Sequence 106, App
29	11	68.8	104	1	US-08-899-575-106	Sequence 106, App
30	11	68.8	104	1	US-08-899-575-106	Sequence 106, App
31	11	68.8	104	5	PCT-US95-08743-106	Sequence 106, App
32	11	68.8	105	1	US-08-276-852-89	Sequence 89, Appl
33	11	68.8	105	1	US-08-899-575-89	Sequence 89, Appl
34	11	68.8	105	1	US-08-899-575-89	Sequence 89, Appl
35	11	68.8	105	5	PCT-US95-08743-89	Sequence 89, Appl
36	11	68.8	106	1	US-08-276-852-83	Sequence 83, Appl
37	11	68.8	106	1	US-08-276-852-85	Sequence 85, Appl
38	11	68.8	106	1	US-08-899-575-83	Sequence 83, Appl
39	11	68.8	106	1	US-08-899-575-85	Sequence 85, Appl
40	11	68.8	106	1	US-08-899-575-83	Sequence 83, Appl
41	11	68.8	106	1	US-08-899-575-85	Sequence 85, Appl
42	11	68.8	106	3	US-09-240-274-159	Sequence 159, App
43	11	68.8	106	3	US-09-240-274-165	Sequence 165, App
44	11	68.8	106	5	PCT-US95-08743-83	Sequence 83, Appl
45	11	68.8	106	5	PCT-US95-08743-85	Sequence 85, Appl

ALIGNMENTS

RESULT 1  
US-08-540-412-123  
; Sequence 123, Application US/08540412  
; Patent No. 5866679  
; GENERAL INFORMATION:  
; APPLICANT: DeFeo-Jones, Deborah  
; APPLICANT: Feng, Dong-Mei  
; APPLICANT: Garsky, Victor M.  
; APPLICANT: Jones, Raymond E.  
; APPLICANT: Oliff, Allen I.  
; TITLE OF INVENTION: NOVEL PEPTIDES  
; NUMBER OF SEQUENCES: 194  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID A. MUTHARD  
; STREET: 126 E. Lincoln Avenue, P.O. BOX 2000  
; CITY: RAHWAY  
; STATE: NEW JERSEY  
; COUNTRY: U.S.A.  
; ZIP: 07065  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: /US/08/540,412  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Muthard, David A.  
; REGISTRATION NUMBER: 35,297  
; REFERENCE/DOCKET NUMBER: 192531C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908)594-3903  
; TELEFAX: (908)594-4720  
; INFORMATION FOR SEQ ID NO: 123:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
US-08-540-412-123

Query Match 58.8%; Score 11; DB 2; Length 11;  
Best Local Similarity 33.3%; Pred. No. 8.9e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
Db 6 YASSSL 11

STREET: 126 E. Lincoln Avenue, P.O. BOX 2000  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: U.S.A.  
ZIP: 07065

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/051,342  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Muthard, David A.  
REGISTRATION NUMBER: 35,297  
REFERENCE/DOCKET NUMBER: 192531CY  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-3903  
TELEFAX: (908)594-4720  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
US-09-051-342-123

Query Match 68.8%; Score 11; DB 3; Length 11;  
Best Local Similarity 33.3%; Pred. No. 8.9e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
Db 6 YASSSL 11

RESULT 4  
US-09-051-342-191  
Sequence 191, Application US/09051342  
Patent No. 6130204  
GENERAL INFORMATION:  
APPLICANT: DeFeo-Jones, Deborah  
APPLICANT: Feng, Dong-Mei  
APPLICANT: Garsky, Victor M.  
APPLICANT: Jones, Raymond E.  
APPLICANT: Oliff, Allen I.  
TITLE OF INVENTION: NOVEL PEPTIDES  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID A. MUTHARD  
STREET: 126 E. Lincoln Avenue, P.O. BOX 2000  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: U.S.A.  
ZIP: 07065

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/051,342  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Muthard, David A.  
REGISTRATION NUMBER: 35,297  
REFERENCE/DOCKET NUMBER: 192531CY  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-3903  
TELEFAX: (908)594-4720  
INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-412-191

Query Match 68.8%; Score 11; DB 2; Length 11;  
Best Local Similarity 33.3%; Pred. No. 8.9e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
Db 6 YASSSL 11

RESULT 3  
US-09-051-342-123  
Sequence 123, Application US/09051342  
Patent No. 6130204  
GENERAL INFORMATION:  
APPLICANT: DeFeo-Jones, Deborah  
APPLICANT: Feng, Dong-Mei  
APPLICANT: Garsky, Victor M.  
APPLICANT: Jones, Raymond E.  
APPLICANT: Oliff, Allen I.  
TITLE OF INVENTION: NOVEL PEPTIDES  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID A. MUTHARD

REFERENCE/DOCKET NUMBER: 192531CY  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-3903  
TELEFAX: (908)594-4720  
INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-051-342-191

Query Match 68.8%; Score 11; DB 3; Length 11;  
Best Local Similarity 33.3%; Pred. No. 8.9e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0;

QY 1 YXXXXL 6  
DB 6 YASSSL 11

RESULT 5  
US-08-468-161-123

; Sequence 123, Application US/08468161  
; Patent No. 6143864

GENERAL INFORMATION:

APPLICANT: DeFeo-Jones, Deborah

APPLICANT: Peng, Dong-Mei

APPLICANT: Garsky, Victor M.

APPLICANT: Jones, Raymond E.

APPLICANT: Oliff, Allen I.

TITLE OF INVENTION: NOVEL PEPTIDES

NUMBER OF SEQUENCES: 146

CORRESPONDENCE ADDRESS:

ADDRESSEE: DAVID A. MUTHARD

STREET: 126 E. Lincoln Avenue, P.O. BOX 2000

CITY: RAHWAY

STATE: NEW JERSEY

COUNTRY: U.S.A.

ZIP: 07065

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,161

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Muthard, David A.

REGISTRATION NUMBER: 35,297

REFERENCE/DOCKET NUMBER: 192531B

TELECOMMUNICATION INFORMATION:

TELEPHONE: (908)594-3903

TELEFAX: (908)594-4720

INFORMATION FOR SEQ ID NO: 123:

SEQUENCE CHARACTERISTICS:

LENGTH: 11 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal

US-08-468-161-123

Query Match 68.8%; Score 11; DB 3; Length 11;  
Best Local Similarity 33.3%; Pred. No. 8.9e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0;

QY 1 YXXXXL 6  
DB 6 YASSSL 11

RESULT 6

US-09-051-759-123

; Sequence 123, Application US/09051759

; Patent No. 6177404

GENERAL INFORMATION:

APPLICANT: Merck & Co., Inc.

APPLICANT: DeFeo-Jones, Deborah

APPLICANT: Garsky, Victor M.

APPLICANT: Jones, Raymond E.

APPLICANT: Oliff, Allen I.

APPLICANT: Scolnick, Edward M.

TITLE OF INVENTION: CONJUGATES USEFUL IN THE TREATMENT OF

TITLE OF INVENTION: BENIGN PROSTATIC HYPERPLASIA

FILE REFERENCE: 19560P

CURRENT APPLICATION NUMBER: US/09/051,759

CURRENT FILING DATE: 1998-08-03

PRIOR APPLICATION NUMBER: 60/005,664

PRIOR FILING DATE: 1995-10-18

PRIOR APPLICATION NUMBER: PCT/US96/16490

PRIOR FILING DATE: 1996-10-15

NUMBER OF SEQ ID NOS: 194

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 123

LENGTH: 11

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: completely synthetic amino acid sequence

US-09-051-759-123

Query Match 68.8%; Score 11; DB 3; Length 11;

Best Local Similarity 33.3%; Pred. No. 8.9e+02;

Matches 2; Conservative 0; Mismatches 4; Indels 0;

Gaps 0;

QY 1 YXXXXL 6  
DB 6 YASSSL 11

RESULT 7

US-09-051-759-191

; Sequence 191, Application US/09051759

; Patent No. 6177404

GENERAL INFORMATION:

APPLICANT: Merck & Co., Inc.

APPLICANT: DeFeo-Jones, Deborah

APPLICANT: Garsky, Victor M.

APPLICANT: Jones, Raymond E.

APPLICANT: Oliff, Allen I.

APPLICANT: Scolnick, Edward M.

TITLE OF INVENTION: CONJUGATES USEFUL IN THE TREATMENT OF

TITLE OF INVENTION: BENIGN PROSTATIC HYPERPLASIA

FILE REFERENCE: 19560P

CURRENT APPLICATION NUMBER: US/09/051,759

CURRENT FILING DATE: 1998-08-03

PRIOR APPLICATION NUMBER: 60/005,664

PRIOR FILING DATE: 1995-10-18

PRIOR APPLICATION NUMBER: PCT/US96/16490

PRIOR FILING DATE: 1996-10-15

NUMBER OF SEQ ID NOS: 194

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 191

LENGTH: 11

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: completely synthetic amino acid sequence

US-09-051-759-191

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/541,779
; FILING DATE: 21-JUN-1990
; SEQ ID NO:5:
; LENGTH: 18
5219837-5

Query Match      68.8%; Score 11; DB 6; Length 18;
Best Local Similarity 33.3%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
Db 12 YSASTL 17

RESULT 10
US-08-905-223-22
; Sequence 22, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; FILING DATE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 1..37
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 5.9
; OTHER INFORMATION: seq LSYASSALSPCLT/AP
;
US-08-905-223-22

Query Match      68.8%; Score 11; DB 3; Length 37;
Best Local Similarity 33.3%; Pred. No. 2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
Db 27 YASSAL 32

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/541,779
; FILING DATE: 21-JUN-1990
; SEQ ID NO:5:
; LENGTH: 18
5219837-5

Query Match      68.8%; Score 11; DB 3; Length 11;
Best Local Similarity 33.3%; Pred. No. 8.9e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
Db 6 YASSSL 11

RESULT 8
PCT-US95-08156-123
; Sequence 123, Application PC/TUS9508156
; GENERAL INFORMATION:
; APPLICANT: DeFeo-Jones, Deborah
; APPLICANT: Feng, Dong-Mei
; APPLICANT: Garsky, Victor M.
; APPLICANT: Jones, Raymond E.
; APPLICANT: Oliff, Allen I.
; TITLE OF INVENTION: NOVEL PEPTIDES
; NUMBER OF SEQUENCES: 146
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID A. MUTHARD
; STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: U.S.A.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08156
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 192531B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3903
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 123:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
;
PCT-US95-08156-123

Query Match      68.8%; Score 11; DB 5; Length 11;
Best Local Similarity 33.3%; Pred. No. 8.9e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
Db 6 YASSSL 11

RESULT 9
5219837-5
; Patent No. 5219837
; APPLICANT: COHEN, JEFFREY A.; GREENE, MARK I.; WILLIAMS,
; WILLIAM V.
; TITLE OF INVENTION: METHOD OF STIMULATING MYELINATION
; OF CELLS
; NUMBER OF SEQUENCES: 9
```



```
RESULT 11
US-09-247-155-22
; Sequence 22, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 22
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo Sapiens
; NAME/KEY: SIGNAL
; LOCATION: 1..37
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.9
; OTHER INFORMATION: seq LSYASSALSPCLT/AP
US-09-663-600A-22
Query Match 68.8%; Score 11; DB 4; Length 37;
Best Local Similarity 33.3%; Pred. No. 2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 YXXXXL 6
Db 27 YASSAL 32
RESULT 13
US-08-812-586-40
; Sequence 40, Application US/08812586
; Patent No. 6048704
; GENERAL INFORMATION:
; APPLICANT: Martin David Tilson
; TITLE OF INVENTION: PURIFIED AND RECOMBINANT ANTIGENIC
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH ABDOMINAL AORTIC ANEURYSM (AAA)
; TITLE OF INVENTION: DISEASE, AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/812,586
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/53862-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-812-586-40
Query Match 68.8%; Score 11; DB 3; Length 40;
Best Local Similarity 33.3%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 YXXXXL 6
Db 15 YAASTL 20
RESULT 14
US-09-535-832A-37
```

```
; Sequence 37, Application US/09535832A
; Patent No. 6537769
; GENERAL INFORMATION:
; APPLICANT: Tilson, Martin David
; TITLE OF INVENTION: Purified and Recombinant Antigenic Proteins Associated
; TITLE OF INVENTION: with Abdominal Aortic Aneurysm (AAA) Disease, and
; TITLE OF INVENTION: Diagnostic and Therapeutic use Thereof
; FILE REFERENCE: 53862-AZ
; CURRENT APPLICATION NUMBER: US/09/535,832A
; CURRENT FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; TYPE: PRT
; LENGTH: 40
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Source:
; OTHER INFORMATION: Synthesized
US-09-535-832A-37

Query Match          68.8%; Score 11; DB 4; Length 40;
Best Local Similarity 33.3%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 YXXXXL 6
Db      15 YAASTL 20

RESULT 15
US-09-107-532A-4990
; Sequence 4990, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4990:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
```

```
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...64
; SEQUENCE DESCRIPTION: SEQ ID NO: 4990:
US-09-107-532A-4990

Query Match          68.8%; Score 11; DB 4; Length 64;
Best Local Similarity 33.3%; Pred. No. 2.9e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 YXXXXL 6
Db      49 YAAATSL 54

Search completed: October 5, 2004, 16:15:36
Job time : 10.4932 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:12:48 ; Search time 33.7534 Seconds  
(without alignments)  
66.737 Million cell updates/sec

Title: US-09-973-473A-21

Perfect score: 16

Sequence: 1.YXXXXLX 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	68.8	17	14	US-10-105-545-8
2	11	68.8	17	15	US-10-120-835-65
3	11	68.8	18	14	US-10-105-545-21
4	11	68.8	18	15	US-10-409-643-54
5	11	68.8	19	15	US-10-120-835-66
6	11	68.8	20	15	US-10-197-945A-2
7	11	68.8	27	12	US-10-424-599-275349
8	11	68.8	28	15	US-10-120-835-67
9	11	68.8	36	12	US-10-424-599-250421
10	11	68.8	37	10	US-09-903-190-22
11	11	68.8	37	14	US-10-319-763-22
12	11	68.8	37	16	US-10-437-963-153132
13	11	68.8	38	15	US-10-437-963-120714
14	11	68.8	41	12	US-10-276-774-2215
15	11	68.8	43	16	US-10-109-048-661

16	11	68.8	44	16	US-10-767-701-61017	Sequence 61017, A
17	11	68.8	45	12	US-10-424-599-208666	Sequence 208666, A
18	11	68.8	45	16	US-10-109-048-869	Sequence 869, App
19	11	68.8	48	12	US-10-424-599-169349	Sequence 169349, A
20	11	68.8	49	16	US-10-437-963-142578	Sequence 142578, A
21	11	68.8	50	12	US-09-978-360A-624	Sequence 624, App
22	11	68.8	51	12	US-10-424-599-154170	Sequence 154170, A
23	11	68.8	52	14	US-10-218-102-329	Sequence 329, App
24	11	68.8	53	12	US-10-424-599-217394	Sequence 217394, A
25	11	68.8	55	12	US-10-424-599-171812	Sequence 171812, A
26	11	68.8	55	12	US-10-424-599-199688	Sequence 199688, A
27	11	68.8	55	16	US-10-437-963-179375	Sequence 179375, A
28	11	68.8	56	16	US-10-767-701-48322	Sequence 48322, A
29	11	68.8	57	16	US-10-437-963-201040	Sequence 201040, A
30	11	68.8	58	16	US-10-437-963-153608	Sequence 153608, A
31	11	68.8	59	9	US-09-864-761-34665	Sequence 34665, A
32	11	68.8	59	16	US-10-437-963-188808	Sequence 188808, A
33	11	68.8	60	12	US-10-424-599-187770	Sequence 187770, A
34	11	68.8	61	9	US-09-764-898-156	Sequence 156, App
35	11	68.8	61	12	US-10-424-599-154747	Sequence 154747, A
36	11	68.8	64	12	US-10-424-599-211017	Sequence 211017, A
37	11	68.8	64	12	US-10-424-599-257942	Sequence 257942, A
38	11	68.8	64	16	US-10-767-701-50784	Sequence 50784, A
39	11	68.8	65	16	US-10-437-963-141623	Sequence 141623, A
40	11	68.8	66	16	US-10-767-701-50448	Sequence 50448, A
41	11	68.8	70	12	US-10-424-599-197254	Sequence 197254, A
42	11	68.8	70	12	US-10-424-599-244085	Sequence 244085, A
43	11	68.8	70	16	US-10-437-963-127594	Sequence 127594, A
44	11	68.8	72	12	US-10-424-599-185136	Sequence 185136, A
45	11	68.8	72	16	US-10-437-963-183123	Sequence 183123, A

#### ALIGNMENTS

#### RESULT 1

US-10-105-545-8  
; Sequence 8, Application US/10105545  
; Publication No. US20030144479A1  
; GENERAL INFORMATION:  
; APPLICANT: Mark, Greene I.  
; APPLICANT: Williams, William V.  
; APPLICANT: Weiner, David B.  
; APPLICANT: Cohen, Jeffrey A.  
; APPLICANT: Kieber-Emmons, Thomas  
; APPLICANT: Williams, Robert M.  
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE COMPOUNDS AND METHODS OF CONSTRUCTING AND USING  
; TITLE OF INVENTION: SAME  
; FILE REFERENCE: 4040/LL492US2  
; CURRENT APPLICATION NUMBER: US/10105545  
; CURRENT FILING DATE: 2002-09-06  
; PRIOR APPLICATION NUMBER: US 08/752,816  
; PRIOR FILING DATE: 1996-11-21  
; PRIOR APPLICATION NUMBER: US 07/940,654  
; PRIOR FILING DATE: 1992-09-03  
; PRIOR APPLICATION NUMBER: US 07/702,833  
; PRIOR FILING DATE: 1991-05-20  
; PRIOR APPLICATION NUMBER: US 07/326,328  
; PRIOR FILING DATE: 1989-03-21  
; PRIOR APPLICATION NUMBER: US 07/074,264  
; PRIOR FILING DATE: 1987-07-16  
; PRIOR APPLICATION NUMBER: US 07/462,542  
; PRIOR FILING DATE: 1990-01-09  
; PRIOR APPLICATION NUMBER: US 07/648,303  
; PRIOR FILING DATE: 1991-01-25  
; PRIOR APPLICATION NUMBER: US 07/685,881  
; PRIOR FILING DATE: 1991-04-15  
; PRIOR APPLICATION NUMBER: US 07/574,391  
; PRIOR FILING DATE: 1990-08-27  
; PRIOR APPLICATION NUMBER: US 07/194,026  
; PRIOR FILING DATE: 1988-05-13  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 32

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-105-545-8

Query Match      68.8%; Score 11; DB 14; Length 17;
Best Local Similarity 33.3%; Pred. No. 2.9e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 YXXXXL 6
DB      11 YSASTL 16

RESULT 2
US-10-120-835-65
; Sequence 65, Application US/10120835
; Publication No. US20040018189A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Mitchell, Steve C.
; APPLICANT: Afar, Daniel E. H.
; APPLICANT: Saffran, Douglas
; APPLICANT: Morrison, Karen J. M.
; APPLICANT: Morrison, Robert K.
; APPLICANT: Ge, Wangmao
; APPLICANT: Jakobovitz, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
; FILE REFERENCE: 51158-20061.00
; CURRENT APPLICATION NUMBER: US/10/120,835
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/282,739
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/286,630
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/300,373
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-120-835-65

Query Match      68.8%; Score 11; DB 15; Length 17;
Best Local Similarity 33.3%; Pred. No. 2.9e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 YXXXXL 6
DB      5 YSTTTL 10

RESULT 3
US-10-105-545-21
; Sequence 21, Application US/10105545
; Publication No. US20030144479A1
; GENERAL INFORMATION:
; APPLICANT: Mark, Greene I.
; APPLICANT: Williams, William V.
; APPLICANT: Weiner, David B.
; APPLICANT: Cohen, Jeffery A.
; APPLICANT: Kieber-Emmons, Thomas
; APPLICANT: Kieber-Emmons, Thomas

; APPLICANT: Williams, Robert M.
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE COMPOUNDS AND METHODS OF CONSTRUCTING AND USING
; TITLE OF INVENTION: SAME
; FILE REFERENCE: 4040/1L492US2
; CURRENT APPLICATION NUMBER: US/10/105,545
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 08/752,816
; PRIOR FILING DATE: 1996-11-21
; PRIOR APPLICATION NUMBER: US 07/940,654
; PRIOR FILING DATE: 1992-09-03
; PRIOR APPLICATION NUMBER: US 07/702,833
; PRIOR FILING DATE: 1991-05-20
; PRIOR APPLICATION NUMBER: US 07/326,328
; PRIOR FILING DATE: 1989-03-21
; PRIOR APPLICATION NUMBER: US 07/074,264
; PRIOR FILING DATE: 1987-07-16
; PRIOR APPLICATION NUMBER: US 07/462,542
; PRIOR FILING DATE: 1990-01-09
; PRIOR APPLICATION NUMBER: US 07/648,303
; PRIOR FILING DATE: 1991-01-25
; PRIOR APPLICATION NUMBER: US 07/685,881
; PRIOR FILING DATE: 1991-04-15
; PRIOR APPLICATION NUMBER: US 07/574,391
; PRIOR FILING DATE: 1990-08-27
; PRIOR APPLICATION NUMBER: US 07/194,026
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-10-105-545-21

Query Match      68.8%; Score 11; DB 14; Length 18;
Best Local Similarity 33.3%; Pred. No. 3.1e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 YXXXXL 6
DB      12 YSASTL 17

RESULT 4
US-10-409-643-54
; Sequence 54, Application US/10409643
; Publication No. US20030235577A1
; GENERAL INFORMATION:
; APPLICANT: Shapiro, Steven
; APPLICANT: Hartzell, William
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR PREVENTING AND TREATING MICROBIAL IN
; FILE REFERENCE: B00801.70281.US
; CURRENT APPLICATION NUMBER: US/10/409,643
; CURRENT FILING DATE: 2003-04-08
; PRIOR APPLICATION NUMBER: US 60/370,649
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-409-643-54

Query Match      68.8%; Score 11; DB 15; Length 18;
Best Local Similarity 33.3%; Pred. No. 3.1e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 YXXXXL 6
DB      1 YXXXXL 6
```

```
Db          9 YSASTL 14

RESULT 5
US-10-120-835-66
; Sequence 66, Application US/10120835
; Publication No. US20040018189A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Mitchell, Steve C.
; APPLICANT: Afar, Daniel E. H.
; APPLICANT: Saffran, Douglas
; APPLICANT: Morrison, Karen J. M.
; APPLICANT: Morrison, Robert K.
; APPLICANT: Ge, Wangmao
; APPLICANT: Jakobovitz, Ava
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
; FILE REFERENCE: ENTITLED 121P2A3 USEFUL IN TREATMENT AND DETECTION OF CANCER
; CURRENT APPLICATION NUMBER: US/10/120,835
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/282,739
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/286,630
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/300,373
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-120-835-66

Query Match          68.8%; Score 11; DB 15; Length 19;
Best Local Similarity 33.3%; Pred. No. 3.2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy          1 YXXXXL 6
Db          6 YSTTTL 11

RESULT 6
US-10-197-945A-2
; Sequence 2, Application US/10197945A
; Publication No. US20040014148A1
; GENERAL INFORMATION:
; APPLICANT: Masuda, Esteban
; APPLICANT: Kinella, Todd M
; APPLICANT: Warner, Justin E
; APPLICANT: Kinoshita, Taisei
; APPLICANT: Bennett, Mark K
; APPLICANT: Anderson, David C
; TITLE OF INVENTION: Methods of Identifying Compounds that Modulate IL-4 Receptor-Mediated Signaling
; FILE REFERENCE: RIGL-013/000S
; CURRENT APPLICATION NUMBER: US/10/197,945A
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Peptide generated by a combinatorial library
US-10-197-945A-2

Db          9 YSASTL 14

Query Match          68.8%; Score 11; DB 15; Length 19;
Best Local Similarity 33.3%; Pred. No. 3.2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy          1 YXXXXL 6
Db          6 YSTTTL 11

RESULT 6
US-10-197-945A-2
; Sequence 2, Application US/10197945A
; Publication No. US20040014148A1
; GENERAL INFORMATION:
; APPLICANT: Masuda, Esteban
; APPLICANT: Kinella, Todd M
; APPLICANT: Warner, Justin E
; APPLICANT: Kinoshita, Taisei
; APPLICANT: Bennett, Mark K
; APPLICANT: Anderson, David C
; TITLE OF INVENTION: Methods of Identifying Compounds that Modulate IL-4 Receptor-Mediated Signaling
; FILE REFERENCE: RIGL-013/000S
; CURRENT APPLICATION NUMBER: US/10/197,945A
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Peptide generated by a combinatorial library
US-10-197-945A-2

Db          9 YSASTL 14

Query Match          68.8%; Score 11; DB 15; Length 20;
Best Local Similarity 33.3%; Pred. No. 3.3e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy          1 YXXXXL 6
Db          2 YTSAAAL 7

RESULT 7
US-10-424-599-275349
; Sequence 275349, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With Soy Nucleic Acid Molecules and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 275349
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_90660C.1.pap
US-10-424-599-275349

Query Match          68.8%; Score 11; DB 12; Length 27;
Best Local Similarity 33.3%; Pred. No. 4.2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy          1 YXXXXL 6
Db          12 YTSSTL 17

RESULT 8
US-10-120-835-67
; Sequence 67, Application US/10120835
; Publication No. US20040018189A1
; GENERAL INFORMATION:
; APPLICANT: Agensys, Inc.
; APPLICANT: Challita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Mitchell, Steve C.
; APPLICANT: Afar, Daniel E. H.
; APPLICANT: Saffran, Douglas
; APPLICANT: Morrison, Karen J. M.
; APPLICANT: Morrison, Robert K.
; APPLICANT: Ge, Wangmao
; APPLICANT: Jakobovitz, Ava
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
; FILE REFERENCE: ENTITLED 121P2A3 USEFUL IN TREATMENT AND DETECTION OF CANCER
; CURRENT APPLICATION NUMBER: US/10/120,835
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/282,739
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/286,630
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/300,373
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67
```

```

; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-120-835-67

Query Match      68.8%; Score 11; DB 15; Length 28;
Best Local Similarity 33.3%; Pred. No. 4.3e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 11 YSTTTL 16

RESULT 9
US-10-424-599-250421
; Sequence 22, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 250421
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT MRT3847_68158C.1.pep
US-10-424-599-250421

Query Match      68.8%; Score 11; DB 12; Length 36;
Best Local Similarity 33.3%; Pred. No. 5.2e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 22 YSTSSL 27

RESULT 10
US-09-903-190-22
; Sequence 22, Application US/09903190
; Publication No. US20030162176A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/903,190
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US/09/247,155A
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/074,121
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/081,563
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/096,116
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/099,273
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 22
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: 1..37
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.9
; OTHER INFORMATION: seq LSYASSALSPCLT/AP
US-10-319-763-22

Query Match      68.8%; Score 11; DB 14; Length 37;
Best Local Similarity 33.3%; Pred. No. 5.3e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 27 YASSAL 32

RESULT 11
US-10-319-763-22
; Sequence 22, Application US/10319763
; Publication No. US20030144490A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS
; FILE REFERENCE: G-031.US04.DIV
; CURRENT APPLICATION NUMBER: US/10/319,763
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/066,677
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/069,957
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/074,121
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/081,563
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/096,116
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/099,273
; PRIOR FILING DATE: 1998-09-04
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: Patent.pm
; SEQ ID NO 22
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: 1..37
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.9
; OTHER INFORMATION: seq LSYASSALSPCLT/AP
US-10-319-763-22

Query Match      68.8%; Score 11; DB 14; Length 37;
Best Local Similarity 33.3%; Pred. No. 5.3e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 27 YASSAL 32

RESULT 12
US-10-437-963-153132
; Sequence 153132, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
```

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 153132
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(41)
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53114C.1.pep
US-10-437-963-153132

Query Match      68.8%; Score 11; DB 16; Length 37;
Best Local Similarity 33.3%; Pred. No. 5.3e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 21 YASASL 26

RESULT 13
US-10-437-963-120714
; Sequence 120714, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 120714
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(41)
; OTHER INFORMATION: Clone ID: PAT_MRT4530_23809C.1.pep
US-10-437-963-120714

Query Match      68.8%; Score 11; DB 16; Length 38;
Best Local Similarity 33.3%; Pred. No. 5.4e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 1 YSTAAL 6

RESULT 14
US-10-276-774-2215
; Sequence 2215, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
```

```
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2215
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(41)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2215

Query Match      68.8%; Score 11; DB 12; Length 41;
Best Local Similarity 33.3%; Pred. No. 5.7e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 10 YTSSTL 15

RESULT 15
US-10-109-048-661
; Sequence 661, Application US/10109048
; Publication No. US20040107461A1
; GENERAL INFORMATION:
; APPLICANT: COMMURI, PADMA
; APPLICANT: KEELING, PETER L.
; APPLICANT: RAMIREZ, NONA
; APPLICANT: MCKEAN, ANGELA
; APPLICANT: GAO, ZHONG
; APPLICANT: GUAN, HANPING
; TITLE OF INVENTION: GLUCAN CHAIN LENGTH DOMAINS
; FILE REFERENCE: 2461-76
; CURRENT APPLICATION NUMBER: US/10/109,048
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: 60/279,720
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 1154
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 661
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(41)
; OTHER INFORMATION: Description of Unknown Organism: Accession No. 2833388
US-10-109-048-661

Query Match      68.8%; Score 11; DB 16; Length 43;
Best Local Similarity 33.3%; Pred. No. 6e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 12 YATAAL 17

Search completed: October 5, 2004, 16:47:02
Job time : 35.7534 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:58:01 ; Search time 8.53425 Seconds  
(without alignments)  
78.899 Million cell updates/sec

Title: US-09-973-473A-21

Perfect score: 16

Sequence: 1 YXXXXLX 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	68.8	15	2 PH1314	Ig heavy chain DJ
2	11	68.8	38	2 D82306	hypothetical prote
3	11	68.8	46	2 E85743	hypothetical prote
4	11	68.8	54	2 J70521	Ig kappa chain V-I
5	11	68.8	61	2 E39741	14K hypothetical t
6	11	68.8	68	1 JN0537	head protein gp3 -
7	11	68.8	68	2 F30970	probable head comp
8	11	68.8	69	1 QVZ7	hypothetical prote
9	11	68.8	69	2 P42516	D-ORF-C protein -
10	11	68.8	71	2 S21526	Ig kappa chain V r
11	11	68.8	76	2 T17673	hypothetical prote
12	11	68.8	81	2 C99872	hypothetical prote
13	11	68.8	84	2 F87306	conserved hypothet
14	11	68.8	86	2 S34086	Ig kappa chain V r
15	11	68.8	86	2 S16840	Ig kappa chain V r
16	11	68.8	87	2 S21523	Ig kappa chain V r
17	11	68.8	87	2 S34084	Ig kappa chain V r
18	11	68.8	87	2 S34083	Ig kappa chain V r
19	11	68.8	88	2 S21528	Ig kappa chain V r
20	11	68.8	88	2 S21524	Ig kappa chain V r
21	11	68.8	88	2 S21525	Ig kappa chain V r
22	11	68.8	88	2 S21520	Ig kappa chain V r
23	11	68.8	88	2 S21522	Ig kappa chain V r
24	11	68.8	88	2 P10261	Ig kappa chain V r
25	11	68.8	88	2 S34104	Ig kappa chain V r
26	11	68.8	90	2 B56273	sakacin A immunity
27	11	68.8	91	2 S17622	Ig kappa chain V r
28	11	68.8	91	2 JQ2204	hypothetical 10.7K
29	11	68.8	93	2 D71874	hypothetical prote

30	11	68.8	94	2 A01955	Ig kappa-B5 chain
31	11	68.8	95	2 PH0863	Ig kappa chain V r
32	11	68.8	95	2 S45324	Ig kappa chain V r
33	11	68.8	95	2 S69898	Ig kappa chain V r
34	11	68.8	98	2 S11062	Ig light chain V r
35	11	68.8	101	2 C28840	Ig kappa chain V r
36	11	68.8	101	2 B28840	Ig kappa chain V r
37	11	68.8	101	2 B37262	Ig kappa chain V r
38	11	68.8	101	2 S44117	Ig kappa chain V-J
39	11	68.8	104	2 F87731	protein W10C8.2 (i
40	11	68.8	105	2 S36266	Ig lambda chain V
41	11	68.8	106	2 P10260	Ig kappa chain V r
42	11	68.8	106	2 P10259	Ig kappa chain V r
43	11	68.8	106	2 P10262	Ig kappa chain V r
44	11	68.8	107	2 S47183	Ig kappa chain - h
45	11	68.8	107	2 S36264	Ig lambda chain V

ALIGNMENTS

RESULT 1

PH1314

Ig heavy chain DJ region (Clone C200-98) - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C/Accession: PH1314

R/Wasserman, R.; Galili, N.; Ito, Y.; Reichard, B.A.; Shane, S.; Rovera, G.

J. Exp. Med. 176, 1577-1581, 1992

A>Title: Predominance of fetal type DJH joining in young children with B precursor lymphoma

A/Reference number: PH1302; MUID:93094761; PMID:1460419

A/Accession: PH1314

A/Molecule type: DNA

A/Residues: 1-15 <WAS>

C/Keywords: heterotetramer; immunoglobulin

Query Match 68.8%; Score 11; DB 2; Length 15;

Best Local Similarity 33.3%; Pred. No. 82;

Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6

Db 2 YSSSL 7

RESULT 2

D82306

hypothetical protein VC0584 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C/Species: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C/Accession: D82306

R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.

l. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; MUID:20406833; PMID:10952301

A/Accession: D82306

A>Status: preliminary

A/Molecule type: DNA

A/Residues: 1-38 <HEI>

A/Cross-references: GB:AE004143; GB:AE003852; NID:G9655003; PTDN:AAF93751.1; GSPDB:GN0013

A/Experimental source: serogroup O1; strain N16961; biotype El Tor

C/Genetics:

A/Gene: VC0584

A/Map position: 1

Query Match 68.8%; Score 11; DB 2; Length 38;

Best Local Similarity 33.3%; Pred. No. 1.7e+02;

Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6

J. Biol. Chem. 266, 7793-7803, 1991  
A;Title: Discovery and sequence analysis of bacterial genes involved in the biogenesis of  
A;Reference number: A39741; UID:91210304; PMID:1850420  
A;Accession: E39741  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-61 <RAM>  
A;Cross-references: GB:M60874; NID:gl52073; PIDN:AAA26195.1; PID:gl52077  
C;Keywords: transmembrane protein

Db 30 YSTTSL 35

RESULT 3  
E85743  
hypotheoretical protein Z2363 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: E85743; B85717  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; UID:21074935; PMID:11206551  
A;Accession: E85743  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-46 <STO>  
A;Cross-references: GB:AE005174; NID:gl2515355; PIDN:AA056409.1; GSPDB:GN00145; UWGP:Z23  
A;Experimental source: strain O157:H7, substrain EDL933  
A;Accession: B85717  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-46 <ST2>  
A;Cross-references: GB:AE005174; NID:gl2515086; PIDN:AA056198.1; GSPDB:GN00145; UWGP:Z21  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z2363; Z2132

Query Match 68.8%; Score 11; DB 2; Length 46;  
Best Local Similarity 33.3%; Pred. No. 1.9e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 13 YTAASL 18

RESULT 4  
JT0521  
Ig kappa chain V-III region (CP1) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 23-Oct-1992 #sequence\_revision 23-Oct-1992 #text\_change 09-May-1997  
C;Accession: JT0521  
R;Anker, R.; Conley, M.E.; Pollok, B.A.  
J. Exp. Med. 169, 2109-2119, 1989  
A;Title: Clonal diversity in the B cell repertoire of patients with X-linked agammaglobu  
A;Reference number: JT0511; UID:89279157; PMID:2786547  
A;Accession: JT0521  
A;Molecule type: mRNA  
A;Residues: 1-54 <ANK>  
A;Note: the sequence shown here is one of eight productive V-D-J mu chain rearrangements  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotetramer; immunoglobulin  
F;1-49/Domain: V region <VRE>  
F;50-54/Domain: J region <JRE>

Query Match 68.8%; Score 11; DB 2; Length 54;  
Best Local Similarity 33.3%; Pred. No. 2.1e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 1 YAASSL 6

RESULT 5  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 6  
JN0537  
head protein gp3 - phage 21  
C;Species: phage 21  
C;Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 23-Jul-1999  
C;Accession: JN0537; C49849  
R;Smith, M.P.; Feiss, M.  
Gene 126, 1-7, 1993  
A;Title: Sequence analysis of the phage 21 genes for prohead assembly and head completor  
A;Reference number: JN0537; UID:93231520; PMID:8472949  
A;Accession: JN0537  
A;Molecule type: DNA  
A;Residues: 1-68 <SM1>  
A;Cross-references: GB:M81255; NID:g215454; PIDN:AAA32341.1; PID:g215457  
R;Smith, M.P.; Feiss, M.  
J. Bacteriol. 175, 2393-2399, 1993  
A;Title: Sites and gene products involved in lambdaoid phage DNA packaging.  
A;Reference number: A49849; UID:93224462; PMID:8468297  
A;Accession: C49849  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-27 <SM2>  
A;Note: sequence extracted from NCBI backbone (NCBIN:129221, NCBIP:129224)  
C;Genetics:  
A;Gene: 3  
C;Superfamily: phage lambda head-to-tail joining protein W  
C;Keywords: head protein

Query Match 68.8%; Score 11; DB 1; Length 68;  
Best Local Similarity 33.3%; Pred. No. 2.6e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 35 YTAASL 40

RESULT 7  
F90970  
probable head completion protein [imported] - Escherichia coli (strain O157:H7, substrain  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-May-2002  
C;Accession: F90970  
R;Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
A;Reference number: A99629; UID:21156231; PMID:11258796  
A;Accession: F90970  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-68 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BA036157.1; PID:gl3362202; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RIMD 0509952

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 8  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 9  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 10  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 11  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 12  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 13  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 14  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 15  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 16  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 17  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 18  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 19  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
| |  
Db 16 YAAAL 21

RESULT 20  
E39741  
14K hypothetical thioredoxin-related cyoX 3' region protein - Bradyrhizobium japonicum  
C;Species: Bradyrhizobium japonicum  
C;Date: 30-Dec-1991 #sequence\_revision 30-Dec-1991 #text\_change 08-Oct-1999  
C;Accession: E39741  
R;Ramseier, T.M.; Winteler, H.V.; Hennecke, H.

Query Match 68.8%; Score 11; DB 2; Length 61;  
Best Local Similarity 33.3%; Pred. No. 2.4e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

C;Genetics:  
A;Gene: EC62734  
A;Superfamily: phage lambda head-to-tail joining protein W

Query Match 68.8%; Score 11; DB 2; Length 68;  
Best Local Similarity 33.3%; Pred. No. 2.6e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 35 YTAASL 40

RESULT 8  
QOV27  
hypothetical protein D-69 - vaccinia virus (strain WR)  
C;Species: vaccinia virus  
C;Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 10-Sep-1999  
C;Accession: A03878  
R;Niles, E.G.; Condit, R.C.; Caro, P.; Davidson, K.; Matusick, L.; Seto, J.  
Virology 153, 96-112, 1986  
A;Title: Nucleotide sequence and genetic map of the 16-kb vaccinia virus HindIII D fragment  
A;Reference number: A01146; PMID:86291159; PMID:3739227  
A;Accession: A03878  
A;Molecule type: DNA  
A;Residues: 1-69 <N1>  
A;Cross-references: GB:M15058  
C;Superfamily: vaccinia virus D-ORF-C protein

Query Match 68.8%; Score 11; DB 1; Length 69;  
Best Local Similarity 33.3%; Pred. No. 2.6e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 8 YASSSL 13

RESULT 9  
F42516  
D-ORF-C protein - vaccinia virus (strain Copenhagen)  
C;Species: vaccinia virus  
A;Note: host Homo sapiens (man)  
C;Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 20-Sep-1999  
C;Accession: F42516  
R;Johnson, G.P.  
submitted to GenBank, June 1990  
A;Reference number: A33172  
A;Accession: F42516  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-69 <JOH>  
C;Superfamily: vaccinia virus D-ORF-C protein

Query Match 68.8%; Score 11; DB 2; Length 69;  
Best Local Similarity 33.3%; Pred. No. 2.6e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 8 YASSSL 13

RESULT 10  
S21526  
Ig kappa chain V region - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 20-Feb-1995 #sequence\_revision 25-Oct-1996 #text\_change 23-Jul-1999  
C;Accession: S34082; S21526  
R;Wagner, S.D.; Luzzatto, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A;Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed  
A;Reference number: S34076; PMID:93170387; PMID:8436174

A;Accession: S34082  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-71 <WA2>  
A;Cross-references: EMBL:X66042; NID:G33318; PIDN:CAA46841.1; PID:G33319  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotetramer; immunoglobulin

Query Match 68.8%; Score 11; DB 2; Length 71;  
Best Local Similarity 33.3%; Pred. No. 2.6e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 42 YAASTL 47

RESULT 11  
TI7673  
hypothetical protein al83L - Chlorella virus PBCV-1  
C;Species: Chlorella virus PBCV-1  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: TI7673  
R;Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A;Reference number: Z18806  
A;Accession: TI7673  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-76 <GRA>  
A;Cross-references: EMBL:U42580; NID:G4028896; PIDN:AAC96551.1  
A;Experimental source: specific host Chlorella strain NC64A  
C;Genetics:  
A;Note: al83L

Query Match 68.8%; Score 11; DB 2; Length 76;  
Best Local Similarity 33.3%; Pred. No. 2.8e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 37 YAAASL 42

RESULT 12  
C89872  
hypothetical protein [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: C89872  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; PMID:21311952; PMID:11418146  
A;Accession: C89872  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-81 <KUR>  
A;Cross-references: GB:BA000018; PID:gl3700838; PIDN:BA842134.1; GSPDB:GN00149  
C;Experimental source: strain N315  
C;Genetics:  
A;Gene: SA0889

Query Match 68.8%; Score 11; DB 2; Length 81;  
Best Local Similarity 33.3%; Pred. No. 2.9e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 27 YTASAL 32

RESULT 13

F87306

conserved hypothetical protein CC0463 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 17-May-2002

C:Accession: F87306

R:NIERMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFET, D.H.; KOLONIN, J.; EMOLAEVA, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87306

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-84 <STO>

A:Cross-references: GB:AE005673; NID:g13421636; PIDN:AAK22450.1; GSPDB:GN00148

C:Genetics:

C:Superfamily: conserved hypothetical protein HI1000

Query Match 68.8%; Score 11; DB 2; Length 84;  
Best Local Similarity 33.3%; Pred. No. 3e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 36 YTAAL 41

RESULT 14

S34086

Ig kappa chain V region - human

C:Species: Homo sapiens (man)

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jan-2000

C:Accession: S34086

R:WAGNER, S.D.; LUZZATTO, L. Eur. J. Immunol. 23, 391-397, 1993

A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed

A:Reference number: S34076; MUID:93170387; PMID:8436174

A:Accession: S34086

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-86 <WAG>

A:Cross-references: EMBL:X67169

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:9-83/Domain: immunoglobulin homology <IMM>

Query Match 68.8%; Score 11; DB 2; Length 86;  
Best Local Similarity 33.3%; Pred. No. 3e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 42 YAASSL 47

RESULT 15

S16840

Ig kappa chain V region - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000

C:Accession: S16840

R:BLAISON, G.; KUNTZ, J.L.; PASQUALI, J.L. Eur. J. Immunol. 21, 1221-1227, 1991

A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid fac

A:Reference number: S16823; MUID:91243737; PMID:1903706

A:Accession: S16840

A>Status: preliminary; translation not shown

A:Molecule type: mRNA

A:Residues: 1-86 <BLA>

A:Cross-references: EMBL:X54838

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:7-81/Domain: immunoglobulin homology <IMM>

Query Match 68.8%; Score 11; DB 2; Length 86;  
Best Local Similarity 33.3%; Pred. No. 3e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
|  
DB 40 YAASSL 45

Search completed: October 5, 2004, 16:13:51  
Job time : 10.5342 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:36 ; Search time 4.89041 Seconds  
(without alignments)  
74.532 Million cell updates/sec

Title: US-09-973-473A-21

Perfect score: 16

Sequence: 1 YXXXLX 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	68.8	61	1	CMND_BRAJA
2	11	68.8	68	1	VGO3_ECOLI6
3	11	68.8	69	1	YVDC_VACCV
4	11	68.8	69	1	YVDC_VACCV
5	11	68.8	84	1	Y463_CAUCR
6	11	68.8	90	1	SATA_LACSK
7	11	68.8	94	1	KV11_RABIT
8	11	68.8	107	1	YNPA_RHIL0
9	11	68.8	108	1	KV1H_HUMAN
10	11	68.8	108	1	KV1K_HUMAN
11	11	68.8	108	1	KV1S_HUMAN
12	11	68.8	110	1	KV01_RABIT
13	11	68.8	111	1	IM9B_CAEEL
14	11	68.8	117	1	KV1I_HUMAN
15	11	68.8	127	1	FUR_SINP7
16	11	68.8	129	1	KV1H_HUMAN
17	11	68.8	130	1	KV5G_MOUSE
18	11	68.8	138	1	PSBR_HORVU
19	11	68.8	139	1	UGR1_MOUSE
20	11	68.8	145	1	THYI_ERIEU
21	11	68.8	147	1	THYI_BOVIN
22	11	68.8	147	1	THYI_SHEEP
23	11	68.8	147	1	THYI_SORAR
24	11	68.8	149	1	HUPT_AZOCH
25	11	68.8	150	1	THYI_PIG
26	11	68.8	189	1	COAT_CMV
27	11	68.8	194	1	INA_FELCA
28	11	68.8	199	1	Y11Q_ECOLI
29	11	68.8	205	1	FGFH_NPVOP
30	11	68.8	209	1	ZFP7_ARATH
31	11	68.8	210	1	CUTI1_PHYCP
32	11	68.8	224	1	YESA_SCHPO
33	11	68.8	225	1	YWBB_BACSU

34	11	68.8	231	1	YKNW_BACSU
35	11	68.8	233	1	SNG4_MOUSE
36	11	68.8	234	1	SNG4_HUMAN
37	11	68.8	237	1	Z198_MOUSE
38	11	68.8	243	1	YCGE_ECOLI
39	11	68.8	261	1	COX3_CHICK
40	11	68.8	267	1	BLOS_PSEAE
41	11	68.8	280	1	PSD_CHRVO
42	11	68.8	285	1	MTAP_CAEEL
43	11	68.8	294	1	DGDR_BURCE
44	11	68.8	298	1	RU1A_YEAST
45	11	68.8	299	1	Y060_MYCPN

ALIGNMENTS

RESULT 1  
CMND\_BRAJA  
ID CMND\_BRAJA STANDARD; PRT; 61 AA.  
AC P30959;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Heme exporter protein D (Cytochrome c-type biogenesis protein cycX).  
GN CYCX OR CMND OR BSR0470.  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=110RIF15;  
RX MEDLINE=91210304; PubMed=1850420;  
RA Rameier T.M., Winteler H.V., Hennecke H.;  
RT "Discovery and sequence analysis of bacterial genes involved in the  
RT biogenesis of c-type cytochromes.";  
RL J. Biol. Chem. 266:7793-7803(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=USDA 110;  
RX MEDLINE=22484998; PubMed=12597275;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,  
RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
RT Bradyrhizobium japonicum USDA110.";  
RL DNA Res. 9:189-197(2002).  
CC -!- FUNCTION: Required for the export of heme to the periplasm for the  
CC biogenesis of C-type cytochromes (Potential).  
CC -!- SIMILARITY: BELONGS TO THE CMND/CYCX/HELD FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; M60874; AAA26195.1; -.  
DR EMBL; AP005936; BAC45735.1; -.  
DR PIR; E39741; E39741.  
DR InterPro; IPR007078; Ccnd.  
DR Pfam; PF04995; Ccnd; 1.  
KW Cytochrome c-type biogenesis; Transport; Transmembrane;  
KW Inner membrane; Complete proteome.  
FT TRANSMEM 11 31  
SQ SEQUENCE 61 AA; 6776 MW; BC6B9852E927370A CRC64;  
Query Match 68.8%; Score 11; DB 1; Length 61;  
Best Local Similarity 33.3%; Pred. No. 1.4e+02;

```
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 YXXXXL 6
Db 16 YAAAL 21

RESULT 2
VG03_ECOL6 STANDARD; PRT; 68 AA.
AC P36271;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2003 (Rel. 42, Last annotation update)
DE Head stabilization protein (Head protein GP3).
GN 3 OR C1570.
OS Escherichia coli O6, and
OC Bacteriophage P21 (Bacteriophage 21).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992, 10711;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Phage P21;
RX MEDLINE=93231520; PubMed=8472949;
RA Smith M.P., Feiss M.;
RT "Sequence analysis of the phage 21 genes for prohead assembly and
head completion.";
RL Gene 126:1-7(1993).
CC -1- SIMILARITY: TO LAMBDA HEAD-TO-TAIL JOINING PROTEIN W.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE016759; AAN80039.1; -.
CC EMBL; M81255; AAA32341.1; -.
CC PIR; JN0537; JN0537.
CC InterPro; IPR004174; gpW.
CC Pfam; PF02831; gpW; 1.
CC ProDom; PD019366; gpW; 1.
CC Complete proteome.
SQ SEQUENCE 68 AA; 7621 MW; 54F7BCD9CEAF2F4 CRC64;

Query Match 68.8%; Score 11; DB 1; Length 68;
Best Local Similarity 33.3%; Pred. No. 1.5e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 YXXXXL 6
Db 35 YTAASL 40

RESULT 3
YVDC_VACCC STANDARD; PRT; 69 AA.
AC P20552;
DT 01-FEB-1991 (Rel. 17, Created)
```

```
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 7.3 kDa protein.
GN D ORF C.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M35027; AAA48101.1; -.
CC PIR; F42516; F42516.
CC Hypothetical protein.
KW SEQUENCE 69 AA; 7347 MW; 73DF495CBD7834BA CRC64;

Query Match 68.8%; Score 11; DB 1; Length 69;
Best Local Similarity 33.3%; Pred. No. 1.5e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 YXXXXL 6
Db 8 YASSSL 13

RESULT 4
YVDC_VACCC STANDARD; PRT; 69 AA.
AC P04304;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Hypothetical 7.3 kDa protein.
GN D ORF C.
OS Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86291159; PubMed=3739227;
RA Niles E.G., Condit R.C., Caro P., Davidson K., Matusick L., Seto J.;
RT "Nucleotide sequence and genetic map of the 16-kb vaccinia virus
HindIII D fragment.";
RL Virology 153:96-112(1986).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
```

```

DR EMBL; M15058; --; NOT_ANNOTATED_CDS.
DR PIR; A03878; QOV27.
KW Hypothetical protein.
SQ SEQUENCE 69 AA; 7321 MW; B9BF494BBD701C18 CRC64;

Query Match      68.8%; Score 11; DB 1; Length 69;
Best Local Similarity 33.3%; Pred. No. 1.9e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 8 YASSSL 13

RESULT 5
ID Y463 CAUCR STANDARD; PRT; 84 AA.
AC Q9AA9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0161 protein CC0463.
GN CC0463
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- SIMILARITY: Belongs to the UPF0161 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AE005719; AAK22450.1; -.
DR PIR; F87306; F87306.
DR TIGR; CC0463; -.
DR HAMAP; MP_00386; -.
DR InterPro; IPR002696; DUF37.
DR Pfam; PF01809; DUF37; 1.
DR ProDom; PD004225; DUF37; 1.
DR TIGRFAMs; TIGR00278; TIGR00278; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 84 AA; 9584 MW; 75E4381C10F04372 CRC64;

Query Match      68.8%; Score 11; DB 1; Length 84;
Best Local Similarity 33.3%; Pred. No. 1.9e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 36 YTAAL 41

RESULT 6
SAIA_LACSK STANDARD; PRT; 90 AA.
ID SAIA_LACSK

```

---

```

AC Q48864;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sakacin A immunity factor.
GN SAIA.
OS Lactobacillus sakei.
OG Plasmid 60 kb.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1599;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LB706;
RX MEDLINE=95238285; PubMed=7721704;
RA Axelsson L., Holick A.;
RT "The genes involved in production of and immunity to sakacin A, a
RT bacteriocin from Lactobacillus sake LB706.";
RL J. Bacteriol. 177:2125-2137(1995).
CC -!- FUNCTION: IMPARTS IMMUNITY TO SAKACIN A TO NATURALLY SENSITIVE
CC HOST STRAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; Z46867; CAA86941.1; -.
DR PIR; B56273; B56273.
KW Bacteriocin immunity; Plasmid.
SQ SEQUENCE 90 AA; 10458 MW; DE9689A2AA18AA1F CRC64;

Query Match      68.8%; Score 11; DB 1; Length 90;
Best Local Similarity 33.3%; Pred. No. 1.9e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 14 YTSTAL 19

RESULT 7
KV11_RABIT STANDARD; PRT; 94 AA.
ID KV11_RABIT
AC P01692;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa-B5 chain V region 2699 (Fragments).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE.
RX MEDLINE=8317897; PubMed=6404296;
RA Ayadi H., Duka S., Paroutaud P., Strosberg A.D.;
RT "Partial amino acid sequence of a rabbit immunoglobulin light chain
RT of allotype b5.";
RL Biochemistry 22:993-998(1983).
CC -!- MISCELLANEOUS: THIS CHAIN IS AN ANTIBODY TO PNEUMOCOCCUS STRAIN
CC III VACCINE.
DR PIR; A01955; A01955.
DR HSSP; P01607; IREI.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003599; IG.
DR SMART; SM00409; IG; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23
FT DOMAIN 24 34
FRAMEWORK-1.
COMPLEMENTARITY-DETERMINING-1.

```

FT DOMAIN 35 49 FRAMEWORK-2.  
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
FT DOMAIN 57 >82 FRAMEWORK-3.  
FT NON CONS 82 83  
FT DOMAIN <83 83 COMPLEMENTARITY-DETERMINING-3.  
FT DOMAIN 84 93 FRAMEWORK-4.  
FT VARIANT 8 8 P -> A.  
FT VARIANT 22 22 N -> K.  
FT NON TER 94 94  
SQ SEQUENCE 94 AA; 9469 MW; 351D977B93252EBC CRC64;

Query Match 68.8%; Score 11; DB 1; Length 94;  
Best Local Similarity 33.3%; Pred. No. 2e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
DB 49 YSASTL 54

RESULT 8  
YNFA\_RHILO STANDARD; PRT; 107 AA.  
ID YNFA\_RHILO STANDARD; PRT; 107 AA.  
AC Q98402; (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE Hypothetical protein ml17841.  
GN MLL7841.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Aemamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
CC -!- SIMILARITY: Belongs to the UPF0060 family.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; AP003012; BAB54221.1; --  
DR HAMAP; MF 00010; --; 1.  
DR InterPro; IPR003844; UPF0060.  
DR Pfam; PF02694; UPF0060; 1.  
DR ProDom; PD015609; UPF0060; 1.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 4 24 POTENTIAL.  
FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 60 80 POTENTIAL.  
FT TRANSMEM 87 107 POTENTIAL.  
SQ SEQUENCE 107 AA; 11591 MW; 0C666D85212C16C3 CRC64;

Query Match 68.8%; Score 11; DB 1; Length 107;  
Best Local Similarity 33.3%; Pred. No. 2.2e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6

DB 6 YTAAL 11

RESULT 9  
KV1H HUMAN STANDARD; PRT; 108 AA.  
ID KV1H HUMAN STANDARD; PRT; 108 AA.  
AC P01603;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-I region Hau.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=71032830; PubMed=4097974;  
RA Watanabe S., Hilschmann N.;  
RT "The primary structure of a monoclonal kappa-type immunoglobulin L-  
chain of subgroup I (Bence-Jones Protein Hau): subdivision within  
subgroups.";  
RL Hoppe-Seyler's Z. Physiol. Chem. 351:1291-1295(1970).  
CC -!- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.  
CC -!- MISCELLANEOUS: This is a Bence-Jones protein.  
DR PIR; A01868; K1HUHU.  
DR HSSP; P80362; 1WTL.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; ig; 1.  
DR SMART; SM00406; Igv; 1.  
DR PROSITE; PS50835; IG LIKE; 1.  
KW Immunoglobulin V region; Bence-Jones protein.  
FT DOMAIN 1 23 FRAMEWORK-1.  
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.  
FT DOMAIN 35 49 FRAMEWORK-2.  
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
FT DOMAIN 57 88 FRAMEWORK-3.  
FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.  
FT DOMAIN 98 107 FRAMEWORK-4.  
FT DISULFID 23 88 BY SIMILARITY.  
FT NON TER 108 108  
SQ SEQUENCE 108 AA; 11671 MW; 08D3A6160D8D0618 CRC64;

Query Match 68.8%; Score 11; DB 1; Length 108;  
Best Local Similarity 33.3%; Pred. No. 2.2e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
DB 49 YAASSL 54

RESULT 10  
KV1K HUMAN STANDARD; PRT; 108 AA.  
ID KV1K HUMAN STANDARD; PRT; 108 AA.  
AC P01603;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-I region Ka.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=76189985; PubMed=818073;  
RA Shinoda T.;



RT "Comparative structural studies on the light chains of human  
 RL immunoglobulin. I. Protein Ka with the Inv(3) allotypic marker.";  
 RL J. Biochem. 77:1277-1296(1975).  
 CC -I- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.  
 CC -I- MISCELLANEOUS: This is a Bence-Jones protein.  
 DR PIR; A01849; KIHUKA.  
 DR HSP; P80362; IWL.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 KW Immunoglobulin V region; Bence-Jones protein.

FT DOMAIN 1 23 FRAMEWORK-1.  
 FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 35 49 FRAMEWORK-2.  
 FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 57 88 FRAMEWORK-3.  
 FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 98 107 FRAMEWORK-4.  
 FT DISULFID 23 88 BY SIMILARITY.  
 FT NON\_TER 108 108  
 SQ SEQUENCE 108 AA; 11900 MW; 768839FBED5A2F4B CRC64;

Query Match 68.8%; Score 11; DB 1; Length 108;  
 Best Local Similarity 33.3%; Pred. No. 2.2e+02;  
 Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
 DB 49 YAASSL 54

## RESULT 11

ID KVIS HUMAN STANDARD; PRT; 108 AA.  
 AC P01611;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-I region Wes.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=81092279; PubMed=6778806;  
 RA Kratzin H., Yang C., Krusche J.U., Hilschmann N.;  
 RT "Preparative separation of the tryptic hydrolysate of a protein by  
 RT high-pressure liquid chromatography. The primary structure of a  
 RT monoclonal L-chain of k-type, subgroup I (Bence-Jones Protein  
 Wes).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 361:1591-1598(1980).  
 CC -I- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.  
 CC -I- MISCELLANEOUS: This is a Bence-Jones protein.  
 DR PIR; A01877; KIHUWS.  
 DR HSP; P80362; IWL.  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0003823; F:antigen binding; NAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 KW Immunoglobulin V region; Bence-Jones protein.

FT DOMAIN 1 23 FRAMEWORK-1.  
 FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 35 49 FRAMEWORK-2.

FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 57 88 FRAMEWORK-3.  
 FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 98 107 FRAMEWORK-4.  
 FT DISULFID 23 88 BY SIMILARITY.  
 FT NON\_TER 108 108  
 SQ SEQUENCE 108 AA; 11608 MW; 782B14A649A60E45 CRC64;

Query Match 68.8%; Score 11; DB 1; Length 108;  
 Best Local Similarity 33.3%; Pred. No. 2.2e+02;  
 Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
 DB 49 YSASSL 54

## RESULT 12

ID KV01 RABIT STANDARD; PRT; 110 AA.  
 AC P01682;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V region 2717.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=74022203; PubMed=4748811;  
 RA Appella E., Roholt O.A., Chersi A., Radzinski G., Pressman D.;  
 RT "Amino acid sequence of the light chain derived from a rabbit anti-p-  
 RT azobenzoyl antibody of restricted heterogeneity.";  
 RL Biochem. Biophys. Res. Commun. 53:1122-1129(1973).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM ANTIBODY TO  
 CC P-AZOBENZONATE AND WAS ISOLATED FROM THE SERUM OF A SINGLE RABBIT.  
 DR PIR; A01945; K4RE27.  
 DR HSP; P80362; IWL.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG; 1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG LIKE; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 23 FRAMEWORK-1.  
 FT DOMAIN 24 36 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 37 51 FRAMEWORK-2.  
 FT DOMAIN 52 58 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 59 90 FRAMEWORK-3.  
 FT DOMAIN 91 99 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 100 109 FRAMEWORK-4.  
 FT SITE 98 98 AT THE HAPTEN COMBINING SITE.  
 FT NON\_TER 110 110  
 SQ SEQUENCE 110 AA; 11367 MW; 8A590BBD5282D107 CRC64;

Query Match 68.8%; Score 11; DB 1; Length 110;  
 Best Local Similarity 33.3%; Pred. No. 2.2e+02;  
 Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6  
 DB 51 YTASSL 56

## RESULT 13

ID IM9B CAEEL STANDARD; PRT; 111 AA.  
 AC Q9Y0V2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

```
DE Mitochondrial import inner membrane translocase subunit Tim9B.
GN TIM9B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Bauer M.F., Brunner M., Hofmann S.;
RT "Cloning and mapping of the Tim10/DDP gene family encoding small zinc
RT finger proteins involved in mitochondrial carrier import.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By
CC similarity).
CC -1- SIMILARITY: Belongs to the Tim8/Tim10 family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF150109; AAD40015.1; -.
DR InterPro: IPR004217; Znf_Tim10/DDP.
DR Pfam; PF02953; zf-Tim10-DDP; 1.
KW Transport; Protein transport; Translocation; Mitochondrion;
KW Inner membrane.
SQ SEQUENCE 111 AA; 12456 MW; C4B9154B15C7439F CRC64;
Query Match 68.8%; Score 11; DB 1; Length 111;
Best Local Similarity 33.3%; Pred. No. 2.2e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 YXXXXL 6
Db 32 YTTSTL 37
RESULT 14
KV11 HUMAN STANDARD; PRT; 117 AA.
ID -KV11 HUMAN STANDARD; PRT; 117 AA.
AC P01601;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-I region HK101 precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bentley D.L., Rabbitts T.H.;
RX MEDLINE=81098966; PubMed=6779204;
RT "Human immunoglobulin variable region genes -- DNA sequences of two V
RT kappa genes and a pseudogene.";
RL Nature 288:730-733(1980).
RN [2]
RP SEQUENCE FROM N.A.
RA Bentley D.L., Rabbitts T.H.;
RX MEDLINE=83129397; PubMed=6402305;
RT "Evolution of immunoglobulin V genes: evidence indicating that
RT recently duplicated human V kappa sequences have diverged by gene
RT conversion.";
RL Cell 32:181-189(1983).
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K01322; AAA58930.1; -.
DR EMBL; K01324; AAA58932.1; -.
DR EMBL; V00558; CAA23824.1; -.
DR PIR; A01881; K1H011.
DR HSSP; P01607; 1REI1.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin V region; Signal.
KW SIGNAL 1 22
FT CHAIN 23 >117 IG KAPPA CHAIN V-I REGION HK101.
FT DOMAIN 23 45 FRAMEWORK-1.
FT DOMAIN 46 56 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 57 71 FRAMEWORK-2.
FT DOMAIN 72 78 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 79 110 FRAMEWORK-3.
FT DOMAIN 111 >117 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 45 110 BY SIMILARITY.
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 12799 MW; D7D0FF3718CBF587 CRC64;
Query Match 68.8%; Score 11; DB 1; Length 117;
Best Local Similarity 33.3%; Pred. No. 2.3e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 YXXXXL 6
Db 71 YAASSL 76
RESULT 15
FUR SYN7 STANDARD; PRT; 127 AA.
ID -FUR SYN7 STANDARD; PRT; 127 AA.
AC Q55344;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ferric uptake regulation protein (ferric uptake regulator).
GN FUR.
OS Synchococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RA Chasemian M., Straus N.A.;
RX MEDLINE=96262718; PubMed=8704986;
RT "Fur regulates the expression of iron-stress genes in the
RT cyanobacterium Synchococcus sp. strain PCC 7942.";
RL Microbiology 142:1469-1476(1996).
CC -1- FUNCTION: ACTS AS A GLOBAL NEGATIVE CONTROLLING ELEMENT, EMPLOYING
CC FE(2+) AS A COFACTOR TO BIND THE OPERATOR OF THE RERESSED GENES
CC REGULATES GENES INVOLVED IN IRON SCAVENGING OR PHOTOSYNTHETIC
CC ELECTRON TRANSPORT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the Fur family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L41065; AAB41546.1; -.
DE Mitochondrial import inner membrane translocase subunit Tim9B.
GN TIM9B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Bauer M.F., Brunner M., Hofmann S.;
RT "Cloning and mapping of the Tim10/DDP gene family encoding small zinc
RT finger proteins involved in mitochondrial carrier import.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By
CC similarity).
CC -1- SIMILARITY: Belongs to the Tim8/Tim10 family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF150109; AAD40015.1; -.
DR InterPro: IPR004217; Znf_Tim10/DDP.
DR Pfam; PF02953; zf-Tim10-DDP; 1.
KW Transport; Protein transport; Translocation; Mitochondrion;
KW Inner membrane.
SQ SEQUENCE 111 AA; 12456 MW; C4B9154B15C7439F CRC64;
Query Match 68.8%; Score 11; DB 1; Length 111;
Best Local Similarity 33.3%; Pred. No. 2.2e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 YXXXXL 6
Db 32 YTTSTL 37
RESULT 14
KV11 HUMAN STANDARD; PRT; 117 AA.
ID -KV11 HUMAN STANDARD; PRT; 117 AA.
AC P01601;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-I region HK101 precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bentley D.L., Rabbitts T.H.;
RX MEDLINE=81098966; PubMed=6779204;
RT "Human immunoglobulin variable region genes -- DNA sequences of two V
RT kappa genes and a pseudogene.";
RL Nature 288:730-733(1980).
RN [2]
RP SEQUENCE FROM N.A.
RA Bentley D.L., Rabbitts T.H.;
RX MEDLINE=83129397; PubMed=6402305;
RT "Evolution of immunoglobulin V genes: evidence indicating that
RT recently duplicated human V kappa sequences have diverged by gene
RT conversion.";
RL Cell 32:181-189(1983).
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K01322; AAA58930.1; -.
DR EMBL; K01324; AAA58932.1; -.
DR EMBL; V00558; CAA23824.1; -.
DR PIR; A01881; K1H011.
DR HSSP; P01607; 1REI1.
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin V region; Signal.
KW SIGNAL 1 22
FT CHAIN 23 >117 IG KAPPA CHAIN V-I REGION HK101.
FT DOMAIN 23 45 FRAMEWORK-1.
FT DOMAIN 46 56 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 57 71 FRAMEWORK-2.
FT DOMAIN 72 78 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 79 110 FRAMEWORK-3.
FT DOMAIN 111 >117 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 45 110 BY SIMILARITY.
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 12799 MW; D7D0FF3718CBF587 CRC64;
```

DR InterPro; IPR002481; FUR.  
DR Pfam; PF01475; FUR; 1.  
DR ProDom; PD002003; FUR; 1.  
KW Transcription regulation; Repressor; DNA-binding; Iron; Zinc.  
FT DOMAIN 73 76 HIS-RICH  
FT METAL 79 79 ZINC (BY SIMILARITY).  
FT METAL 82 82 ZINC (BY SIMILARITY).  
SQ SEQUENCE 127 AA; 14451 MW; 71586B16BCA69426 CRC64;

Query Match 68.8%; Score 11; DB 1; Length 127;  
Best Local Similarity 33.3%; Pred. No. 2.5e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 YXXXXL 6  
|  
Db 3 YTAASL 8

Search completed: October 5, 2004, 16:07:26  
Job time : 6.89041 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:56:41 ; Search time 27.2329 Seconds  
(without alignments)  
81.102 Million cell updates/sec

Title: US-09-973-473a-21  
Perfect score: 16  
Sequence: 1 YXXXXLX 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	68.8	38	16	Q9KUD8 vibrio chol
2	11	68.8	49	2	Q9AFM3
3	11	68.8	61	2	Q8VMA9
4	11	68.8	67	16	Q7UD50
5	11	68.8	68	16	Q8XC12
6	11	68.8	68	16	Q8FIH6
7	11	68.8	69	2	Q9AMS1
8	11	68.8	76	10	Q7X1J7
9	11	68.8	76	12	Q84503
10	11	68.8	81	16	Q99V57
11	11	68.8	84	10	Q8RU88
12	11	68.8	84	11	Q9R0W7
13	11	68.8	84	11	Q9R0W8
14	11	68.8	86	16	Q7VCF2
15	11	68.8	89	2	Q9AH63
16	11	68.8	89	2	Q8GH33

17	11	68.8	90	16	Q7U884
18	11	68.8	93	2	Q9Z4G4
19	11	68.8	93	16	Q9ZKN3
20	11	68.8	96	8	Q8LX75
21	11	68.8	97	5	Q8IF35
22	11	68.8	98	10	Q65086
23	11	68.8	101	2	Q9F320
24	11	68.8	103	2	Q9AH57
25	11	68.8	103	2	Q9AH48
26	11	68.8	106	16	Q8ZQ58
27	11	68.8	107	4	Q9UL81
28	11	68.8	107	4	Q96SA9
29	11	68.8	108	4	Q9UL70
30	11	68.8	108	4	Q9UL77
31	11	68.8	108	4	Q9UL79
32	11	68.8	109	17	Q9YC72
33	11	68.8	111	5	Q8IQU5
34	11	68.8	111	13	Q7T3X9
35	11	68.8	112	2	Q9AH67
36	11	68.8	112	2	Q9AH53
37	11	68.8	113	16	Q87WR0
38	11	68.8	114	16	Q8X2U9
39	11	68.8	117	4	Q13863
40	11	68.8	117	17	Q9YAK5
41	11	68.8	119	2	Q9ALE1
42	11	68.8	119	2	Q9ALE2
43	11	68.8	120	2	Q9ALGO
44	11	68.8	120	2	Q9ALD8
45	11	68.8	120	2	Q9ALF6

ALIGNMENTS

RESULT 1

Q9KUD8 ID Q9KUD8 PRELIMINARY; PRT; 38 AA.

AC Q9KUD8;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Hypothetical protein VC0584.  
GN VC0584.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
RT cholerae.",  
RL Nature 406:477-483(2000).  
DR EMBL; AE004143; AAF93751.1; -.  
DR PIR; D82306; D82306.  
DR TIGR; VC0584; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 38 AA; 4133 MW; BA86FA5823A8899 CRC64;

Query Match 68.8%; Score 11; DB 16; Length 38;  
Best Local Similarity 33.3%; Pred. NO. 8.6e+02;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6

```
Db 30 YSTSL 35

RESULT 2
Q9AFM3 PRELIMINARY; PRT; 49 AA.
ID Q9AFM3
AC Q9AFM3
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Orf, hypothetical.
GN YACB.
OS Shigella flexneri.
OG Plasmid virulence plasmid pWR501.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21189246; PubMed=11292750;
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,
RA Blattner F.R.;
RT "Complete DNA sequence and analysis of the large virulence plasmid of
RT Shigella flexneri."
RL Infect.Immun. 69:3271-3285 (2001).
DR EMBL; AF348706; AAK18546.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Plasmid.
SQ SEQUENCE 49 AA; 5174 MW; 81FEC09F34312CA0 CRC64;

Query Match 68.8%; Score 11; DB 2; Length 49;
Best Local Similarity 33.3%; Pred. No. 1.1e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 12 YSASSL 17

RESULT 3
Q8VMA9 PRELIMINARY; PRT; 61 AA.
ID Q8VMA9
AC Q8VMA9
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE DAP decarboxylase (fragment).
GN LYSA.
OS Rhizobium etli.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=29449;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CE3;
RA Ferraioli S.;
RT "Auxotrophic mutant strains of Rhizobium etli reveal new nodule
RT development phenotypes.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ422134; CAD19513.1; -.
FT NON_TER 1
FT NON_TER 61
SQ SEQUENCE 61 AA; 6955 MW; B94868D845DAAE9A CRC64;

Query Match 68.8%; Score 11; DB 2; Length 61;
Best Local Similarity 33.3%; Pred. No. 1.3e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 31 YSTATL 36
```

```
RESULT 4
Q7UDB0 PRELIMINARY; PRT; 67 AA.
ID Q7UDB0
AC Q7UDB0
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative head completion protein gp3.
GN S0738.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect.Immun. 71:2775-2786 (2003).
DR EMBL; AE016980; AAP16212.1; -.
SQ SEQUENCE 67 AA; 7517 MW; 118F09D14BB19426 CRC64;

Query Match 68.8%; Score 11; DB 16; Length 67;
Best Local Similarity 33.3%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 35 YTAASL 40

RESULT 5
Q8XC12 PRELIMINARY; PRT; 68 AA.
ID Q8XC12
AC Q8XC12
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative head completion protein of prophage CP-9330 (Putative DNA
DE packaging protein of prophage CP-933R).
GN Z2363 OR Z2132 OR ECS2734.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis M.W., Lim A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22 (2001).
```

```
DR EMBL; AE005349; AAG56198.1; ALT INIT.
DR EMBL; AE005368; AAG56409.1; ALT_INIT.
DR EMBL; AP002559; BAB36157.1; -.
DR PIR; E85743; E85743.
DR GO; GO:0019067; P: viral assembly, maturation, egress, and rel. .; IEA.
DR InterPro; IPR004174; gpW.
DR Pfam; PF02831; gpW; 1.
DR ProDom; PD019366; gpW; 1.
KW Complete proteome.
SQ SEQUENCE 68 AA; 7674 MW; 54E2A9D9CEA1F2F4 CRC64;

Query Match 68.8%; Score 11; DB 16; Length 68;
Best Local Similarity 33.3%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 35 YTAASL 40

RESULT 6
ID Q8FIH6 PRELIMINARY; PRT; 68 AA.
AC Q8FIH6;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Putative DNA packaging protein of prophage.
GN C1446
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016759; AAN79915.1; -.
DR GO; GO:0019067; P: viral assembly, maturation, egress, and rel. .; IEA.
DR InterPro; IPR004174; gpW.
DR Pfam; PF02831; gpW; 1.
KW Complete proteome.
SQ SEQUENCE 68 AA; 7700 MW; 428E79D9CEA1F2F4 CRC64;

Query Match 68.8%; Score 11; DB 16; Length 68;
Best Local Similarity 33.3%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 35 YTAASL 40

RESULT 7
ID Q9AMS1 PRELIMINARY; PRT; 69 AA.
AC Q9AMS1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ID909.
GN ID909.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.

DR EMBL; AE005349; AAG56198.1; ALT INIT.
DR EMBL; AE005368; AAG56409.1; ALT_INIT.
DR EMBL; AP002559; BAB36157.1; -.
DR PIR; E85743; E85743.
DR GO; GO:0019067; P: viral assembly, maturation, egress, and rel. .; IEA.
DR InterPro; IPR004174; gpW.
DR Pfam; PF02831; gpW; 1.
DR ProDom; PD019366; gpW; 1.
KW Complete proteome.
SQ SEQUENCE 68 AA; 7674 MW; 54E2A9D9CEA1F2F4 CRC64;

Query Match 68.8%; Score 11; DB 16; Length 68;
Best Local Similarity 33.3%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 35 YTAASL 40

RESULT 8
ID Q7XIJ7 PRELIMINARY; PRT; 76 AA.
AC Q7XIJ7;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein OJ1773_H01.120.
GN OJ1773_H01.120.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OJ1773_H01."
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003932; BAC79669.1; -.
KW Hypothetical protein.
SQ SEQUENCE 76 AA; 8168 MW; 9B8998D17A6AFB0A CRC64;

Query Match 68.8%; Score 11; DB 10; Length 76;
Best Local Similarity 33.3%; Pred. No. 1.6e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YXXXXL 6
DB 33 YSAATL 38

RESULT 9
ID Q84503 PRELIMINARY; PRT; 76 AA.
AC Q84503;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE A183L protein.
GN A183L.
OS Paramesidium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96187795; PubMed=8614977;
RA Lu Z., Li Y., Que Q., Kutish G.F., Rock D.L., Van Etten J.L.;
RT "Analysis of 94 kb of the chlorella virus PCV-1 330-kb genome: map
```

RT positions 98 to 192.";  
RL Virology 216:102-123 (1996).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20013326; PubMed=10544099;  
RA Kaiser A., Vollmert M., Tholl B., Graves M.V., Gurnon J.R., Xing W.,  
RA Lisec A.D., Nickerson K.W., Van Etten J.L.;  
RT "Chlorella virus PBCV-1 encodes a functional homosperrmidine  
RT synthase.";  
RL Virology 263:254-262 (1999).  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20478034; PubMed=11021991;  
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;  
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus  
RT PBCV-1.";  
RL Virology 276:27-36 (2000).  
[4]  
RP SEQUENCE FROM N.A.  
RA Van Etten J.L.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RA Van Etten J.L.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
[6]  
RP SEQUENCE FROM N.A.  
RA Van Etten J.L.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
[7]  
RP SEQUENCE FROM N.A.  
RA Graves M.V., Van Etten J.L.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
[8]  
RP SEQUENCE FROM N.A.  
RA Graves M.V., Van Etten J.L.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
[9]  
RP SEQUENCE FROM N.A.  
RA Gurnon J.R., Graves M.V., Van Etten J.L.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U42580; AAC96551.1; -.  
DR PIR; T17673; T17673.  
SQ SEQUENCE 76 AA; 8237 MW; 4CF5E97A7B46477 CRC64;  
  
Query Match 68.8%; Score 11; DB 12; Length 76;  
Best Local Similarity 33.3%; Pred. No. 1.6e+03;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 YXXXXL 6  
Db 37 YAAASL 42  
  
RESULT 10  
Q99V57 PRELIMINARY; PRT; 81 AA.  
AC Q99V57;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein SAV1036 (Hypothetical protein MW0917).  
GN SAV1036 OR SA0889 OR MW0917.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
OS Staphylococcus aureus (strain N315), and  
OS Staphylococcus aureus (strain MW21).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=158878, 158879, 196620;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Mu50, and N315;  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaiko C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,  
RT aureus.";  
RL Lancet 357:1225-1240 (2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MW2;  
RX MEDLINE=22040717; PubMed=12044378;  
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
RA Yamamoto K., Hiramatsu K.;  
RT "Genome and virulence determinants of high virulence community-  
RT acquired MRSA.";  
RL Lancet 359:1819-1827 (2002).  
DR EMBL; AP003361; BAB57198.1; -.  
DR EMBL; AP003132; BAB42134.1; -.  
DR EMBL; AP004825; BAB94782.1; -.  
DR PIR; C89872; C89872.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 81 AA; 9275 MW; 32EFABC71AD58974 CRC64;  
  
Query Match 68.8%; Score 11; DB 16; Length 81;  
Best Local Similarity 33.3%; Pred. No. 1.7e+03;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 YXXXXL 6  
Db 27 YTASAL 32  
  
RESULT 11  
Q8RUAB PRELIMINARY; PRT; 84 AA.  
AC Q8RUAB;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE OJ1014\_G12.4 protein (B1110C07.28 protein).  
GN OJ1014\_G12.4 OR B1110C07.28  
OS Oryza sativa (japonica cultivar-group). Embryophyta; Tracheophyta;  
OC Eukaryota; Viridiplantae; Streptophyta; Poaceae;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=39947;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC  
RT clone:OJ1014\_G12.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC  
RT clone:B1110C07.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP003372; BAB89070.1; -.  
DR EMBL; AP003902; BAB90627.1; -.  
DR Gramene; Q8RUAB; -.  
SQ SEQUENCE 84 AA; 9564 MW; EALFEI3308091077 CRC64;  
  
Query Match 68.8%; Score 11; DB 10; Length 84;  
Best Local Similarity 33.3%; Pred. No. 1.7e+03;  
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 YXXXXL 6



```
Db 13 YSTSL 18

RESULT 12
Q9ROW7 PRELIMINARY; PRT; 84 AA.
AC Q9ROW7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE Selenium liver binding protein (AP-56).
GN LP82.
OS Mus spretus (Western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Spain;
RA Dragani T.A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X95322; CAA64629.1; -.
SQ SEQUENCE 84 AA; 9406 MW; B17B9CA3DC2A761B CRC64;

Query Match 68.8%; Score 11; DB 11; Length 84;
Best Local Similarity 33.3%; Pred. No. 1.7e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 12 YATSL 17

RESULT 13
Q9ROW8 PRELIMINARY; PRT; 84 AA.
AC Q9ROW8;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE Selenium liver binding protein SP56 (SLP-56).
GN LP8B1.
OS Mus spretus (Western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Spain;
RA Dragani T.A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X95321; CAA64628.1; -.
FT VARIANT 20 20 A -> D.
FT VARIANT 35 35 M -> V.
FT VARIANT 39 39 I -> V.
SQ SEQUENCE 84 AA; 9408 MW; 77CCFCA4B8535EB9 CRC64;

Query Match 68.8%; Score 11; DB 11; Length 84;
Best Local Similarity 33.3%; Pred. No. 1.7e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 12 YATSL 17

RESULT 14
Q7VCF2 PRELIMINARY; PRT; 86 AA.
AC Q7VCF2;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
```

---

```
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Predicted protein.
GN PRO0788.
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=1219;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SARG / CCMF 1375 / SS120;
RX MEDLINE=22810154; PubMed=12917486;
RA Dufréne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,
RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,
RA Makarova K.S., Ostrowski M., Oztas S., Robert C., Rogozin I.B.,
RA Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,
RA Wolf Y.I., Hess W.R.;
RT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,
RT a nearly minimal oxyphototrophic genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).
DR EMBL; AB017163; AAP99832.1; -.
KW Complete proteome.
SQ SEQUENCE 86 AA; 9935 MW; 48A3B08D724BA2E3 CRC64;

Query Match 68.8%; Score 11; DB 16; Length 86;
Best Local Similarity 33.3%; Pred. No. 1.7e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 21 YSTATL 26

RESULT 15
Q9AH63 PRELIMINARY; PRT; 89 AA.
AC Q9AH63;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 24, Last annotation update)
DE PqIA (Fragment).
GN PqIA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMB;
RX MEDLINE=21116988; PubMed=11179344;
RA Kahler C.M., Blum E., Miller Y.K., Ryan D., Popovic T., Stephens D.S.;
RT "exl, an exchangeable genetic island in Neisseria meningitidis.",
RL Infect. Immun. 69:1687-1696(2001).
DR EMBL; AF319530; AAK08028.1; -.
DR InterPro; IPR007498; PqIA.
DR Pfam; PF04403; PqIA; 1.
FT NON TER 89
SQ SEQUENCE 89 AA; 9624 MW; 133931BD96B135C8 CRC64;

Query Match 68.8%; Score 11; DB 2; Length 89;
Best Local Similarity 33.3%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YXXXXL 6
Db 69 YAAASL 74

Search completed: October 5, 2004, 16:12:19
Job time : 31.2329 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:06 ; Search time 56.5068 Seconds  
(without alignments)  
55.003 Million cell updates/sec

Title: US-09-973-473A-22  
Perfect score: 22  
Sequence: 1 XXXXXFFXX 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	59.1	6	ADA89445	Ada89445 Hypersens
2	13	59.1	7	AAE31849	Aae31849 Androgen
3	13	59.1	9	AAR89362	Aar89362 Immunogen
4	13	59.1	9	AAV05025	Aav05025 Tumour an
5	13	59.1	9	AAV98477	Aav98477 HLA class
6	13	59.1	9	AAV98475	Aav98475 HLA class
7	13	59.1	9	AAV98476	Aav98476 HLA class
8	13	59.1	9	AAV98474	Aav98474 HLA class
9	13	59.1	9	AAV984528	Aav984528 Human leu
10	13	59.1	9	AAU06288	Aau06288 Human leu
11	13	59.1	9	AAV98280	Aav98280 Human leu
12	13	59.1	9	ABP25096	Abp25096 Human MHC
13	13	59.1	9	ABP25090	Abp25090 Human MHC
14	13	59.1	9	ABP25092	Abp25092 Human MHC
15	13	59.1	9	ABP11509	Abp11509 HLA class
16	13	59.1	9	ABP25094	Abp25094 Human MHC
17	13	59.1	9	ABP25091	Abp25091 Human MHC
18	13	59.1	9	ABP25095	Abp25095 Human MHC
19	13	59.1	9	AAV98376	Aav98376 Human leu
20	13	59.1	9	AAJ03825	Aaj03825 Hepatitis
21	13	59.1	9	AAJ00066	Aaj00066 Hepatitis
22	13	59.1	9	AAJ03826	Aaj03826 Hepatitis
23	13	59.1	9	AAJ03820	Aaj03820 Hepatitis
24	13	59.1	9	AAJ03824	Aaj03824 Hepatitis
25	13	59.1	9	AAJ00068	Aaj00068 Hepatitis

26	13	59.1	9	4	AAJ03821	Aaj03821 Hepatitis
27	13	59.1	9	4	AAJ00067	Aaj00067 Hepatitis
28	13	59.1	9	4	AAJ00069	Aaj00069 Hepatitis
29	13	59.1	9	4	AAJ03822	Aaj03822 Hepatitis
30	13	59.1	9	5	ABJ09694	Abj09694 Hepatitis
31	13	59.1	9	5	ABJ09689	Abj09689 Hepatitis
32	13	59.1	9	5	ABJ09692	Abj09692 Hepatitis
33	13	59.1	9	5	ABJ05800	Abj05800 Hepatitis
34	13	59.1	9	5	ABJ05798	Abj05798 Hepatitis
35	13	59.1	9	5	ABJ09693	Abj09693 Hepatitis
36	13	59.1	9	5	ABJ09690	Abj09690 Hepatitis
37	13	59.1	9	5	ABJ09688	Abj09688 Hepatitis
38	13	59.1	9	5	ABJ05801	Abj05801 Hepatitis
39	13	59.1	9	5	ABJ05799	Abj05799 Hepatitis
40	13	59.1	9	6	ABJ37976	Abj37976 Human cyt
41	13	59.1	9	7	ADC82768	Adc82768 CDR regio
42	13	59.1	12	2	AAW80391	Aaw80391 Peptide e
43	13	59.1	12	3	AAV58436	Aav58436 Staphyloc
44	13	59.1	12	4	AAV69509	Aav69509 Staphyloc
45	13	59.1	13	2	AAW58716	Aaw58716 Tryptic 4

ALIGNMENTS

RESULT 1  
ADA89445  
ID ADA89445 standard; peptide; 6 AA.  
XX  
AC ADA89445;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Hypersensitive response elicitor receptor AtHrBPlp peptide SEQ ID NO:70.  
XX  
KW receptor; plant; plant pathogen hypersensitive response elicitor;  
KW hypersensitive response elicitor; disease resistance;  
KW enhancing plant growth; controlling insect; stress tolerance.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO2003054211-A2.  
XX  
PD 03-JUL-2003.  
XX  
PF 31-OCT-2002; 2002WO-US035252.  
XX  
PR 31-OCT-2001; 2001US-0335776P.  
PR 17-JUN-2002; 2002US-00174209.  
XX  
PA (EDEN-) EDEN BIOSCIENCE CORP.  
XX  
PI Song X, Bariola PA, Linderoth NA, Fan H, Wei Z;  
XX WPI; 2003-559153/52.  
DR  
XX New isolated protein as a receptor in plants for plant pathogen  
PT hypersensitive response elicitors, useful for identifying agents that  
PT impart disease resistance, enhance plant growth, control insects and/or  
PT impart stress tolerance.  
XX Claim 12; Page 56; 104pp; English.  
XX The present invention describes an isolated protein (I) serving as a  
CC receptor in plants for plant pathogen hypersensitive response elicitors.  
CC Also described: (1) an isolated nucleic acid encoding (I); (2) an  
CC antisense nucleic acid molecule to the nucleic acid of (1); (3) an  
CC expression vector containing the nucleic acid of (1) heterologous to the  
CC nucleic acid of (1) or the DNA molecule of (2); (5) a transgenic plant  
CC transformed with the nucleic acid of (1) or the DNA molecule of (2); (6)  
CC identifying agents targeting plant cells, comprising forming a reaction  
CC mixture having (1) or the host cell of (4) and a candidate agent,

CC evaluating the reaction mixture for binding between (1) or the protein  
CC produced by the host cell and the candidate agent, and identifying  
CC candidate compounds which bind to the proteins in the reaction mixture as  
CC plant cell targeting agents; (7) enhancing plant receptivity to treatment  
CC with hypersensitive response elicitors, comprising providing a transgenic  
CC plant or plant seed transformed with the nucleic acid of (1); and (8)  
CC imparting disease resistance, enhancing growth, controlling insects,  
CC and/or imparting stress resistance to plants, comprising providing a  
CC transgenic plant or plant seed transformed with a DNA construct effective  
CC to silence expression of a nucleic acid molecule of (1), or transformed  
CC with the nucleic acid molecule of (1). The methods and compositions of  
CC the invention are useful for identifying agents targeting plant cells to  
CC enhance a plant's receptivity to treatment with a hypersensitive response  
CC elicitor. The hypersensitive response elicitor treatment includes  
CC imparting disease resistance, enhancing plant growth, controlling insects  
CC and/or imparting stress tolerance. The present sequence represents an  
CC Arabidopsis thaliana hypersensitive response elicitor receptor AtHrP1p  
CC peptide, which is used in an example from the present invention.

XX  
SQ Sequence 6 AA;  
Query Match 59.1%; Score 13; DB 6; Length 6;  
Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
Db 1 YSSAF 5

RESULT 2  
AAE31849  
ID AAE31849 standard; peptide; 7 AA.  
XX  
AC AAE31849;  
DT 07-MAR-2003 (first entry)  
XX  
DE Androgen receptor binding peptide #100.  
XX  
XX Androgen receptor; androgen-associated disorder; prostate cancer; acne;  
KW benign prostatic hypertrophy; hirsutism; androgen insensitivity syndrome;  
KW male pattern baldness; Stein-Leventhal syndrome; infertility; cytostatic;  
KW X-linked spinal bulbar muscular atrophy; antiseborrheic; dermatological;  
KW depilatory; androgen receptor binding peptide.  
XX  
OS Unidentified.  
XX  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 1  
FT /note= "Linked to Xa-Y1; Where Xa is independently a  
FT direct bond or a peptidic structure comprising from about  
FT 1-25 amino acid residues and Y1 is hydrogen, alkyl or  
FT acyl"  
FT  
FT Misc-difference 7  
FT /note= "Linked to Xb-Y2; Where Xb is independently a  
FT direct bond or a peptidic structure comprising from about  
FT 1-25 amino acid residues and Y2 is -OH, amino or  
FT monosubstituted or disubstituted amino"

XX  
PN WO200272612-A2.  
XX  
XX 19-SEP-2002.  
XX  
XX 12-MAR-2002; 2002WO-US007487.  
XX  
XX 12-MAR-2001; 2001US-0275240P.  
PR 28-JAN-2002; 2002US-0352399P.  
XX  
XX (PRAE-) PRAECIS PHARM INC.  
PA  
XX Joyal JL, Mueller J, Oza VB, Findeis MA;  
PI  
XX

DR WPI; 2003-067363/06.  
XX  
XX New peptide modulators of androgen receptor, useful for treating androgen  
PT -associated disorder, e.g. prostate cancer, particularly hormonally  
PT refractive prostate cancer, colon cancer, lung cancer, acne, or  
PT hirsutism.  
XX  
XX Claim 29; Page 36; 68pp; English.  
PS  
XX The present invention relates to novel peptide modulators of androgen  
CC receptor. The peptides of the invention are useful for treating androgen-  
CC associated disorders such as prostate cancer, particularly hormonally  
CC refractive prostate cancer, colon cancer, lung cancer, benign prostatic  
CC hypertrophy, acne, hirsutism, male pattern baldness, Stein-Leventhal  
CC syndrome, androgen insensitivity syndrome, infertility, endometrial  
CC cancer and X-linked spinal bulbar muscular atrophy. The present sequence  
CC is an androgen receptor binding peptide  
XX  
SQ Sequence 7 AA;  
Query Match 59.1%; Score 13; DB 6; Length 7;  
Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
Db 1 YAAAF 5

RESULT 3  
AAR89362  
ID AAR89362 standard; peptide; 9 AA.  
XX  
AC AAR89362;  
DT 18-SEP-1996 (first entry)  
XX  
DE Immunogenic peptide, based on B35 consensus peptide.  
XX  
XX Immunogenic peptide; supermotif; HLA molecule; CTL response; therapeutic;  
KW diagnostic; cancer; viral infection; hepatitis B; hepatitis C.  
XX  
XX Synthetic.  
XX WO9603140-A1.  
PN  
XX 08-FEB-1996.  
PD  
XX 21-JUL-1995; 95WO-US009234.  
PP  
XX 21-JUL-1994; 94US-00278634.  
PR 23-NOV-1994; 94US-00344824.  
PR 30-MAY-1995; 95US-00452843.  
XX  
XX (CVTE-) CYTEL CORP.  
PA  
XX Sette A, Sidney J;  
FI  
XX WPI; 1996-116784/12.  
DR  
XX  
XX Compsn. comprising immunogenic peptide with supermotif allowing more than  
PT one HLA mol. to bind - used to induce CTL response in patient and for in  
PT vivo and ex vivo therapeutic and diagnostic applications.  
XX  
XX Claim 2; Page 26; 32pp; English.  
PS  
XX The sequences given in AAR89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises an  
CC immunogenic peptide of 9-10 residues with a supermotif which allows  
CC binding of more than one HLA molecule. It pref. comprises two conserved  
CC residues, a first at the 2nd position from the N- terminal is Pro, and a  
CC 2nd at the C-terminal is Met. These peptides are used to induce a CTL  
CC response in a patient. They are also useful in compositions for in vivo

CC and ex vivo therapeutic and diagnostic applications, e.g. the treatment of  
CC cancer and viral infections, e.g. hepatitis B and C  
XX  
SQ Sequence 9 AA;

Query Match 59.1%; Score 13; DB 2; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
| |  
Db 5 YAAAF 9

RESULT 4  
ID AAY05025  
XX AAY05025 standard; peptide; 9 AA.

AC AAY05025;

DT 16-JUN-1999 (first entry)

DE Tumour antigen antibody light chain CDR3 clone F3.

XX Tumour antigen; antibody; CDR; complementarity determining region;  
KW binding molecule identification; tumour-specific binding polypeptide;  
KW cancer therapy; light chain.

XX Homo sapiens.

XX WO9906834-A2.

XX 11-FEB-1999.

PF 04-AUG-1998; 98WO-US016280.

PR 04-AUG-1997; 97US-00905825.

PA (IXSY-) IXSYS INC.

PI Watkins JD, Huse WD, Wu H;

XX WPI; 1999-153951/13.

DR N-PSDB; AAX28202.

XX Identifying binding molecules for ligands, particularly tumour antigens -  
PT by selectively immobilising a population of binding molecules to a solid  
PT support and screening for binding to two or more ligands.

XX Claim 15; Page 57; 80pp; English.

XX This sequence represents a light chain complementarity determining region  
CC (CDR) from a tumour antigen specific antibody. The invention relates to a  
CC method for identifying a binding molecule having selective affinity for a  
CC ligand comprising: (a) selectively immobilising a diverse population of  
CC binding molecules to a solid support; (b) simultaneously contacting the  
CC diverse population immobilised on the solid support with 2 or more  
CC ligands; and (c) determining at least one binding molecule which  
CC selectively binds to one or more of the ligands. The method allows for  
CC the rapid and efficient methods for the identification of binding  
CC molecules which exhibit selective affinity for one or more ligands of  
CC interest. They are used particularly for identifying tumour-specific  
CC binding polypeptides which can be used as targeting agents for cancer  
CC therapy that minimises impact on non-tumour tissues

XX Sequence 9 AA;

Query Match 59.1%; Score 13; DB 2; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
| |

Db 3 YSSTP 7

RESULT 5

AAB98477  
ID AAB98477 standard; protein; 9 AA.

XX AAB98477;

DT 22-AUG-2001 (first entry)

DE HLA class I standard peptide binding affinity B\*5401.

XX Human papillomavirus; human leukocyte antigen; HLA; immune response; HPV;  
KW epitope; T cell; identification; vaccine; infection; genital wart;  
KW neoplastic growth; antiviral.

XX Homo sapiens.

XX WO200141799-A1.

PD 14-JUN-2001.

PF 11-DEC-2000; 2000WO-US033549.

XX 10-DEC-1999; 99US-0172705P.

PR 15-AUG-2000; 2000US-00641528.

XX (EPIM-) EPIMUNE INC.

XX Sette A, Sidney J, Southwood S, Chesnut R, Celis E, Grey HM;

XX WPI; 2001-381497/40.

XX An isolated human papilloma virus (HPV) epitope, useful in vaccines for  
PT treating HPV infections.

XX Disclosure; Page 98; 756pp; English.

XX The present invention describes an isolated prepared human papillomavirus  
CC (HPV) epitope (I). (I) has antiviral activity, and can be used in vaccine  
CC production. Peptides and corresponding nucleic acid compositions from the  
CC present invention are useful for stimulating an immune response to HPV by  
CC stimulating the production of CTL or HTL responses, specifically in the  
CC treatment or prophylaxis of HPV infection, in persons who have not  
CC manifested symptoms e.g. genital warts or neoplastic growth. The peptides  
CC can also be used in a tetramer staining assay to assess peripheral blood  
CC mononuclear cells for the presence of antigen-specific CTLs following  
CC exposure to a pathogen or immunogen, and as reagents to evaluate immune  
CC recall responses or evaluate the efficacy of a vaccine. The vaccine  
CC compositions are useful for removing warts or treating HPV infections.  
CC The epitopes for inclusion in an epitope-base vaccine may be selected  
CC from conserved regions of viral or tumour-associated antigens, which  
CC reduces the likelihood of escape mutants, also immunosuppressive epitopes  
CC that may be present in whole antigens can be avoided with the use of  
CC epitope-base vaccines. An additional advantage is the ability to combine  
CC selected epitopes (CTL and HTL) and to modify the composition of the  
CC epitopes achieving enhanced immunogenicity, the major benefit of the  
CC vaccine is that is safe and efficacious. AAB98477 to AAB98477 represent  
CC polypeptide sequences used in the exemplification of the present  
XX invention

XX Sequence 9 AA;

Query Match 59.1%; Score 13; DB 4; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
| |

Db 5 YAAAF 9

```

RESULT 6
AAB98475
ID AAB98475 standard; protein; 9 AA.
XX
AC AAB98475;
XX
DT 22-AUG-2001 (first entry)
XX
DE HLA class I standard peptide binding affinity B51.
XX
KW Human papillomavirus; human leukocyte antigen; HLA; immune response; HPV;
KW epitope; T cell; identification; vaccine; infection; genital wart;
KW neoplastic growth; antiviral.
XX
OS Homo sapiens.
XX
PN WO200141799-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033549.
XX
PR 10-DEC-1999; 99US-0172705P.
XX
PR 15-AUG-2000; 2000US-00641528.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Chesnut R, Celis E, Grey HM;
XX
PI WPI; 2001-381497/40.
XX
PT An isolated human papilloma virus (HPV) epitope, useful in vaccines for
PT treating HPV infections.
XX
PS Disclosure; Page 98; 756pp; English.
XX
CC The present invention describes an isolated prepared human papillomavirus
CC (HPV) epitope (I). (I) has antiviral activity, and can be used in vaccine
CC production. Peptides and corresponding nucleic acid compositions from the
CC present invention are useful for stimulating an immune response to HPV by
CC stimulating the production of CTL or HTL responses, specifically in the
CC treatment or prophylaxis of HPV infection, in persons who have not
CC manifested symptoms e.g. genital warts or neoplastic growth. The peptides
CC can also be used in a tetramer staining assay to assess peripheral blood
CC mononuclear cells for the presence of antigen-specific CTLs following
CC exposure to a pathogen or immunogen, and as reagents to evaluate immune
CC recall responses or evaluate the efficacy of a vaccine. The vaccine
CC compositions are useful for removing warts or treating HPV infections.
CC The epitopes for inclusion in an epitope-base vaccine may be selected
CC from conserved regions of viral or tumour-associated antigens, which
CC reduces the likelihood of escape mutants, also immunosuppressive epitopes
CC that may be present in whole antigens can be avoided with the use of
CC epitope-base vaccines. An additional advantage is the ability to combine
CC selected epitopes (CTL and HTL) and to modify the composition of the
CC epitopes achieving enhanced immunogenicity, the major benefit of the
CC vaccine is that is safe and efficacious. AAB98391 to AAB98477 represent
CC polypeptide sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 9 AA;
Query Match 59.1%; Score 13; DB 4; Length 9;
Best Local Similarity 40.0%; Pred. NO. 1.3e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 YXXXF 9
DB 5 YAAAF 9
RESULT 7
AAB98476
ID AAB98476 standard; protein; 9 AA.

```

```

XX
AC AAB98476;
XX
DT 22-AUG-2001 (first entry)
XX
DE HLA class I standard peptide binding affinity B*5301.
XX
KW Human papillomavirus; human leukocyte antigen; HLA; immune response; HPV;
KW epitope; T cell; identification; vaccine; infection; genital wart;
KW neoplastic growth; antiviral.
XX
OS Homo sapiens.
XX
PN WO200141799-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033549.
XX
PR 10-DEC-1999; 99US-0172705P.
XX
PR 15-AUG-2000; 2000US-00641528.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Chesnut R, Celis E, Grey HM;
XX
PI WPI; 2001-381497/40.
XX
PT An isolated human papilloma virus (HPV) epitope, useful in vaccines for
PT treating HPV infections.
XX
PS Disclosure; Page 98; 756pp; English.
XX
CC The present invention describes an isolated prepared human papillomavirus
CC (HPV) epitope (I). (I) has antiviral activity, and can be used in vaccine
CC production. Peptides and corresponding nucleic acid compositions from the
CC present invention are useful for stimulating an immune response to HPV by
CC stimulating the production of CTL or HTL responses, specifically in the
CC treatment or prophylaxis of HPV infection, in persons who have not
CC manifested symptoms e.g. genital warts or neoplastic growth. The peptides
CC can also be used in a tetramer staining assay to assess peripheral blood
CC mononuclear cells for the presence of antigen-specific CTLs following
CC exposure to a pathogen or immunogen, and as reagents to evaluate immune
CC recall responses or evaluate the efficacy of a vaccine. The vaccine
CC compositions are useful for removing warts or treating HPV infections.
CC The epitopes for inclusion in an epitope-base vaccine may be selected
CC from conserved regions of viral or tumour-associated antigens, which
CC reduces the likelihood of escape mutants, also immunosuppressive epitopes
CC that may be present in whole antigens can be avoided with the use of
CC epitope-base vaccines. An additional advantage is the ability to combine
CC selected epitopes (CTL and HTL) and to modify the composition of the
CC epitopes achieving enhanced immunogenicity, the major benefit of the
CC vaccine is that is safe and efficacious. AAB98391 to AAB98477 represent
CC polypeptide sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 9 AA;
Query Match 59.1%; Score 13; DB 4; Length 9;
Best Local Similarity 40.0%; Pred. NO. 1.3e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 YXXXF 9
DB 5 YAAAF 9
RESULT 8
AAB98474
ID AAB98474 standard; protein; 9 AA.
XX
AC AAB98474;
XX

```

DT 22-AUG-2001 (first entry)  
 XX HLA class I standard peptide binding affinity B\*3510.  
 DE Human papillomavirus; human leukocyte antigen; HLA; immune response; HPV;  
 XX epitope; T cell; identification; vaccine; infection; genital wart;  
 KW neoplastic growth; antiviral.  
 XX Homo sapiens.  
 OS WO200141799-A1.  
 XX PD 14-JUN-2001.  
 XX PF 11-DEC-2000; 2000WO-US033549.  
 XX PR 10-DEC-1999; 99US-0172705P.  
 XX PR 15-AUG-2000; 2000US-00641528.  
 XX PA (EPIM-) EPIMUNE INC.  
 XX PI Sette A, Sidney J, Southwood S, Chesnut R, Celis E, Grey HM;  
 XX WPI; 2001-381497/40.  
 DR An isolated human papilloma virus (HPV) epitope, useful in vaccines for  
 PT treating HPV infections.  
 XX PS Disclosure; Page 98; 756pp; English.  
 XX The present invention describes an isolated prepared human papillomavirus  
 CC (HPV) epitope (I). (I) has antiviral activity, and can be used in vaccine  
 CC production. Peptides and corresponding nucleic acid compositions from the  
 CC present invention are useful for stimulating an immune response to HPV by  
 CC stimulating the production of CTL or HTL responses, specifically in the  
 CC treatment or prophylaxis of HPV infection, in persons who have not  
 CC manifested symptoms e.g. genital warts or neoplastic growth. The peptides  
 CC can also be used in a tetramer staining assay to assess peripheral blood  
 CC mononuclear cells for the presence of antigen-specific CTLs following  
 CC exposure to a pathogen or immunogen, and as reagents to evaluate immune  
 CC recall responses or evaluate the efficacy of a vaccine. The vaccine  
 CC compositions are useful for removing warts or treating HPV infections.  
 CC The epitopes for inclusion in an epitope-base vaccine may be selected  
 CC from conserved regions of viral or tumour-associated antigens, which  
 CC reduces the likelihood of escape mutants, also immunosuppressive epitopes  
 CC that may be present in whole antigens can be avoided with the use of  
 CC epitope-base vaccines. An additional advantage is the ability to combine  
 CC selected epitopes (CTL and HTL) and to modify the composition of the  
 CC epitopes achieving enhanced immunogenicity, the major benefit of the  
 CC vaccine is that is safe and efficacious. AAB98391 to AAB98477 represent  
 CC polypeptide sequences used in the exemplification of the present  
 CC invention  
 XX SQ Sequence 9 AA;  
 Query Match 59.1%; Score 13; DB 4; Length 9;  
 Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 YXXXP 9  
 DB 5 YAAAF 9  
 RESULT 9  
 AAG84528  
 ID AAG84528 standard; peptide; 9 AA.  
 XX AAG84528;  
 AC AAG84528;  
 XX 10-SEP-2001 (first entry)  
 DT Human leukocyte antigen (HLA) class I binding peptide A\*3501.  
 XX

XX Human; human leukocyte antigen; HLA epitope; cytotoxic T lymphocyte; CTL;  
 KW MAGE2; MAGE3; melanoma antigen gene; immune response; vaccine; cancer;  
 KW cytostatic; immunostimulant.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200142267-A1.  
 XX PD 14-JUN-2001.  
 XX PF 11-DEC-2000; 2000WO-US033545.  
 XX PR 10-DEC-1999; 99US-00458298.  
 XX PA (EPIM-) EPIMUNE INC.  
 XX PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;  
 XX Keogh E;  
 XX WPI; 2001-375002/39.  
 DR An isolated prepared MAGE2/3 epitope (I) for use in pharmaceuticals for  
 PT the treatment and prevention of cancer.  
 XX PS Disclosure; Page 83; 171pp; English.  
 XX The present invention describes MAGE2/3 epitopes (I). Also described are:  
 CC (1) a clonal cytotoxic T lymphocyte (CTL) that is cultured in vivo and  
 CC binds to a complex of (I); (2) a peptide (II) comprising (I) and a second  
 CC epitope and has less than 50 contiguous amino acids; (3) a vaccine  
 CC composition comprising (II), a unit dose of a peptide with at least 50  
 CC contiguous amino acids with 100% identity to the native peptide sequence  
 CC of MAGE2/3, and a pharmaceutical excipient; (4) an isolated nucleic acid  
 CC encoding (I); and (5) an isolated nucleic acid encoding (II). (I) has  
 CC cytostatic activity, and can be used in vaccines and as an  
 CC immunostimulant. A vaccine of (3) is useful for the treatment and  
 CC prevention of cancer. (I) is useful for monitoring or evaluating an  
 CC immune response by incubating a T-lymphocyte sample from a patient with  
 CC (I) that binds to an human leukocyte antigen (HLA) allele present in the  
 CC patient and detecting the presence of the T-lymphocyte that binds to the  
 CC peptide. The vaccine allows the opportunity to combine epitopes derived  
 CC from multiple tumour-associated molecules reducing the likelihood of  
 CC tumour escape due to antigen loss. AAG84515 to AAG84909 and AAB99725  
 CC represent amino acid sequences used in the exemplification of the present  
 CC invention  
 XX SQ Sequence 9 AA;  
 Query Match 59.1%; Score 13; DB 4; Length 9;  
 Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 YXXXP 9  
 DB 5 YAAAF 9  
 RESULT 10  
 AAU06288  
 ID AAU06288 standard; peptide; 9 AA.  
 XX AAU06288;  
 AC AAU06288;  
 XX 24-OCT-2001 (first entry)  
 DT Human Leukocyte Antigen (HLA) Class I standard supermotif peptide #9.  
 DE Prostate cancer-associated antigen; supermotif; human leukocyte antigen;  
 XX HLA; cytotoxic T lymphocyte; helper T lymphocyte; HLA superfamily; CTL;  
 KW immunogenicity; immunosuppression; HTL.  
 XX

```

OS Homo sapiens.
PN WO200145728-A2.
XX
XX
PD 28-JUN-2001.
XX
XX 20-DEC-2000; 2000WO-US035516.
XX
XX 21-DEC-1999; 99US-01711312P.
PR 07-AUG-2000; 2000US-00633364.
XX
XX (EPIM-) EPIMUNE INC.
PA
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX WPI; 2001-398311/42.
XX
XX Tumor antigen-associated group-based vaccines useful for vaccinating
PT against prostate cancer.
XX
XX Example 1; Page 85; 252pp; English.
XX
XX The sequences represent prostate cancer-associated antigens and derived
CC motif or supermotif epitopes. The peptide epitopes are included in
CC prostate cancer vaccine compositions due to their ability to bind to
CC human leukocyte antigen (HLA) molecules, which recognise the motifs.
CC Peptides with a high binding affinity are further tested for their
CC ability to induce a cytotoxic T lymphocyte (CTL) or a helper T lymphocyte
CC (HTL) response. Supermotif-bearing peptides may also be tested for their
CC binding affinity to multiple alleles within the HLA superfamily. The
CC vaccine compositions can be modified, for example, to enhance
CC immunogenicity, to avoid the inclusion of immunosuppressive groups, or to
CC alter the immune response to suit the target disease. These group-based
CC vaccines allow the focus of an immune response to multiple selected
CC antigens from the same pathogen. Variability among the immune responses
CC of patients can therefore be alleviated by the inclusion of groups from
CC multiple antigens in a vaccine
XX
XX Sequence 9 AA;
SQ
Query Match 59.1%; Score 13; DB 4; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.3e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 YXXXF 9
Db 5 YAAAF 9
RESULT 11
ID AAG88280 standard; peptide; 9 AA.
XX
XX AAG88280;
XX
XX 11-SEP-2001 (first entry)
XX
XX Human leukocyte antigen (HLA) class I binding peptide A*3501.
XX
XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX WO200141787-A1.
PN
XX 14-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US033591.
XX
XX
PR 10-DEC-1999; 99US-00458299.
XX
XX (EPIM-) EPIMUNE INC.
XX
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX WPI; 2001-374995/39.
XX
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer.
XX
XX Disclosure; Page 83; 199pp; English.
XX
XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising (I)
CC and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and
CC immunostimulant activities, and can be used in vaccines. (I), (II) and
CC (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention
XX
XX Sequence 9 AA;
SQ
Query Match 59.1%; Score 13; DB 4; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.3e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 YXXXF 9
Db 5 YAAAF 9
RESULT 12
ID ABP25096 standard; peptide; 9 AA.
XX
XX ABP25096;
XX
XX 15-JUL-2002 (first entry)
XX
XX Human MHC peptide binding assay peptide #23.
XX
XX HIV-1; human immunodeficiency virus; env; pol; gag; nef; vpr; vpu;
KW vif; tat; cytotoxic T lymphocyte; CTL; immune response; epitope; antigen;
KW vaccine; HIV infection; immunisation; virucide.
XX
XX Homo sapiens.
OS
XX WO200124810-A1.
PN
XX 12-APR-2001.
XX
XX 05-OCT-2000; 2000WO-US027765.
XX

```





```

PI Baker DM, Cellis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-354887/37.
XX
XX Vaccine compositions comprising human immunodeficiency virus-1 (HIV-1)
PT peptide groups, useful for vaccinating against HIV-1.
XX
PS Example 1; Page 416; 448pp; English.
XX
CC The present invention describes a composition (I) comprising a prepared
CC human immunodeficiency virus-1 (HIV-1) group comprising an amino acid
CC sequence selected from 51 defined amino acid sequences (ABL25347 to
CC ABP25397). (I) has virucide activity and can be used in vaccines. (I) may
CC be used for immunising subjects against HIV-1 infections. The use of
CC group-based vaccines has several advantages over traditional vaccines,
CC particularly when compared to the use of whole antigens in vaccine
CC compositions. There is evidence that the immune response to whole
CC antigens is directed largely toward variable regions of the antigen,
CC allowing for immune escape due to mutations. The groups for inclusion in
CC an group-based vaccine may be selected from conserved regions of viral or
CC tumour-associated antigens, which therefore reduces the likelihood of
CC escape mutants. Furthermore, immunosuppressive groups that may be present
CC in whole antigens can be avoided with the use of group-based vaccines. An
CC additional advantage of an group-based vaccine approach is the ability to
CC combine selected groups (CTL and HTL), and further, to modify the
CC composition of the groups, achieving, for example, enhanced
CC immunogenicity. Accordingly, the immune response can be modulated, as
CC appropriate, for the target disease. Similar engineering of the response
CC is not possible with traditional approaches. ABP11501 to ABP25412
CC represent peptide sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 9 AA;
Query Match 59.1%; Score 13; DB 4; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.3e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 YXXXF 9
Db 5 YAAAF 9
RESULT 15
ABP11509
ID ABP11509 standard; peptide; 9 AA.
XX
AC ABP11509;
XX
DT 15-JUL-2002 (first entry)
XX
DE HLA class I binding peptide A*3501.
XX
KW HIV; HIV-1; human immunodeficiency virus; env; pol; gag; nef; vpr; vpu;
KW vif; tat; cytotoxic T lymphocyte; CTL; immune response; epitope; antigen;
KW vaccine; HIV infection; immunisation; virucide.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200124810-A1.
XX
XX 12-APR-2001.
XX
XX 05-OCT-2000; 2000WO-US027766.
XX
XX 05-OCT-1999; 99US-00412863.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Cellis E, Kubo RT, Grey HM;
XX

```

---

```

DR WPI; 2001-354887/37.
XX
XX Vaccine compositions comprising human immunodeficiency virus-1 (HIV-1)
PT peptide groups, useful for vaccinating against HIV-1.
XX
PS Example 1; Page 102; 448pp; English.
XX
CC The present invention describes a composition (I) comprising a prepared
CC human immunodeficiency virus-1 (HIV-1) group comprising an amino acid
CC sequence selected from 51 defined amino acid sequences (ABL25347 to
CC ABP25397). (I) has virucide activity and can be used in vaccines. (I) may
CC be used for immunising subjects against HIV-1 infections. The use of
CC group-based vaccines has several advantages over traditional vaccines,
CC particularly when compared to the use of whole antigens in vaccine
CC compositions. There is evidence that the immune response to whole
CC antigens is directed largely toward variable regions of the antigen,
CC allowing for immune escape due to mutations. The groups for inclusion in
CC an group-based vaccine may be selected from conserved regions of viral or
CC tumour-associated antigens, which therefore reduces the likelihood of
CC escape mutants. Furthermore, immunosuppressive groups that may be present
CC in whole antigens can be avoided with the use of group-based vaccines. An
CC additional advantage of an group-based vaccine approach is the ability to
CC combine selected groups (CTL and HTL), and further, to modify the
CC composition of the groups, achieving, for example, enhanced
CC immunogenicity. Accordingly, the immune response can be modulated, as
CC appropriate, for the target disease. Similar engineering of the response
CC is not possible with traditional approaches. ABP11501 to ABP25412
CC represent peptide sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 9 AA;
Query Match 59.1%; Score 13; DB 4; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.3e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5 YXXXF 9
Db 5 YAAAF 9
Search completed: October 5, 2004, 16:06:32
Job time : 60.5068 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:00:17 ; Search time 14.9178 Seconds  
(without alignments)  
38.068 Million cell updates/sec

Title: US-09-973-473a-22  
Perfect score: 22  
Sequence: 1 XXXXXFXFX 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA+  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	59.1	9	4	US-09-543-608A-50
2	13	59.1	12	2	US-08-811-492-153
3	13	59.1	12	3	US-08-293-728-3
4	13	59.1	12	3	US-09-421-868-3
5	13	59.1	14	1	US-07-841-997A-16
6	13	59.1	14	1	US-08-290-301-16
7	13	59.1	14	4	US-09-013-598-16
8	13	59.1	15	2	US-08-031-538-45
9	13	59.1	15	2	US-08-031-538-56
10	13	59.1	18	1	US-08-401-512-64
11	13	59.1	18	3	US-09-226-012-99
12	13	59.1	28	4	US-09-017-689A-4
13	13	59.1	29	2	US-08-620-151-73
14	13	59.1	29	4	US-09-708-906-7
15	13	59.1	29	4	US-09-708-906-8
16	13	59.1	31	1	US-08-190-802A-239
17	13	59.1	31	3	US-08-477-346-239
18	13	59.1	31	4	US-08-473-089-239
19	13	59.1	31	4	US-08-487-072A-239
20	13	59.1	32	3	US-09-253-396A-214
21	13	59.1	32	3	US-09-708-906-9
22	13	59.1	33	3	US-08-256-104-3
23	13	59.1	37	1	US-08-486-013-16
24	13	59.1	37	2	US-08-482-279-16
25	13	59.1	37	2	US-08-342-268-16
26	13	59.1	37	3	US-09-015-968-16
27	13	59.1	37	4	US-09-397-386-16

28	13	59.1	43	2	US-08-609-046A-2	Sequence 2, Appli
29	13	59.1	43	2	US-08-609-046A-4	Sequence 4, Appli
30	13	59.1	43	3	US-09-158-477-2	Sequence 2, Appli
31	13	59.1	43	3	US-09-158-477-4	Sequence 4, Appli
32	13	59.1	43	4	US-08-858-207A-421	Sequence 421, App
33	13	59.1	45	4	US-08-963-851-32	Sequence 32, Appli
34	13	59.1	56	4	US-09-205-258-1193	Sequence 1193, Ap
35	13	59.1	64	3	US-08-303-861-16	Sequence 16, Appli
36	13	59.1	65	4	US-09-107-532A-5828	Sequence 5828, Ap
37	13	59.1	68	4	US-09-107-532A-4746	Sequence 4746, Ap
38	13	59.1	69	4	US-09-252-991A-18048	Sequence 18048, A
39	13	59.1	83	4	US-09-149-476-346	Sequence 346, App
40	13	59.1	96	1	US-08-486-013-21	Sequence 21, Appli
41	13	59.1	96	2	US-08-482-279-21	Sequence 21, Appli
42	13	59.1	96	2	US-08-342-268-21	Sequence 21, Appli
43	13	59.1	96	3	US-09-015-968-21	Sequence 21, Appli
44	13	59.1	96	4	US-09-397-386-21	Sequence 21, Appli
45	13	59.1	97	4	US-09-439-554-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1  
US-09-543-608A-50  
; Sequence 50, Application US/09543608A  
; Patent No. 6602510  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John D.  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa A.  
; APPLICANT: Cheenut, Robert  
; APPLICANT: Epimmune Inc.  
; TITLE OF INVENTION: HLA Class I A2 Tumor Associated Antigen  
; FILE OF INVENTION: Peptides and Vaccine Compositions  
; FILE REFERENCE: 018623-015710US  
; CURRENT APPLICATION NUMBER: US/09/543,608A  
; CURRENT FILING DATE: 2002-04-05  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 50  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Standard Peptide 1021.05  
US-09-543-608A-50

Query Match 59.1%; Score 13; DB 4; Length 9;  
Best Local Similarity 40.0%; Pred. No. 2.7e+05;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
| |  
Db 5 YAAAF 9

RESULT 2  
US-08-811-492-153  
; Sequence 153, Application US/08811492  
; Patent No. 5834247  
; GENERAL INFORMATION:  
; APPLICANT: COMB, DONALD G.  
; APPLICANT: PERLER, FRANCINE B.  
; APPLICANT: JACK, WILLIAM E.  
; APPLICANT: XU, MING-QUN  
; APPLICANT: HODGES, ROBERT A.  
; APPLICANT: NOREN, CHRISTOPHER J.  
; APPLICANT: CHONG, SHAORONG S.C.  
; APPLICANT: ADAM, ERIC

```

; APPLICANT: SOUTHWORTH, MAURICE
; TITLE OF INVENTION: MODIFIED PROTEINS, METHODS OF THEIR
; TITLE OF INVENTION: PRODUCTION AND METHODS FOR PURIFICATION OF TARGET
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GREGORY D. WILLIAMS; NEW ENGLAND BIOLABS, INC.
; STREET: 32 TOZER ROAD
; CITY: BEVERLY
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC\ DOS\MS\ DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,492
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/496,247
; FILING DATE: 28-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/146,885
; FILING DATE: 03-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/004,139
; FILING DATE: 09-DEC-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Gregory D
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-036C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-927-5054
; TELEFAX: 509-927-1705
; TELEX:
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-811-492-153

Query Match 59.1%; Score 13; DB 2; Length 12;
Best Local Similarity 40.0%; Pred. No. 4.7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
| |
Db 2 YASTF 6

RESULT 3
US-08-293-728-3
; Sequence 3, Application US/08293728D
; Patent No. 6008341
; GENERAL INFORMATION:
; APPLICANT: Foster, Timothy J.
; APPLICANT: McDevitt, Damien L.
; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene
; FILE REFERENCE: 05344.105011
; CURRENT APPLICATION NUMBER: US/08/293,728D
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/293,728
; FILING DATE: 1994-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-421-868-3
Query Match 59.1%; Score 13; DB 3; Length 12;
Best Local Similarity 40.0%; Pred. No. 4.7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
| |
Db 5 YTSAF 9

RESULT 4
US-09-421-868-3
; Sequence 3, Application US/09421868
; Patent No. 6177084
; GENERAL INFORMATION:
; APPLICANT: Foster, Timothy J.
; APPLICANT: McDevitt, Damien L.
; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene
; FILE REFERENCE: 05344.105011
; CURRENT APPLICATION NUMBER: US/09/421,868
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/293,728
; FILING DATE: 1994-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-421-868-3
Query Match 59.1%; Score 13; DB 3; Length 12;
Best Local Similarity 40.0%; Pred. No. 4.7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
| |
Db 5 YTSAF 9

RESULT 5
US-07-841-997A-16
; Sequence 16, Application US/07841997A
; Patent No. 5422254
; GENERAL INFORMATION:
; APPLICANT: Lonsborough, John
; APPLICANT: Vuorio, Outi
; TITLE OF INVENTION: A method to increase the trehalose content
; TITLE OF INVENTION: of organisms by transforming them with the
; TITLE OF INVENTION: structural genes for the short and long chains
; TITLE OF INVENTION: yeast trehalose synthase.
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alko Ltd.
; STREET: PO Box 350
; CITY: Helsinki
; STATE: -
; COUNTRY: Finland
; ZIP: SF-00101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WP5.1 file exported as DOS text file
; CURRENT APPLICATION DATA:

```

```

; CURRENT FILING DATE: 1994-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-08-293-728-3
Query Match 59.1%; Score 13; DB 3; Length 12;
Best Local Similarity 40.0%; Pred. No. 4.7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
| |
Db 5 YTSAF 9

RESULT 4
US-09-421-868-3
; Sequence 3, Application US/09421868
; Patent No. 6177084
; GENERAL INFORMATION:
; APPLICANT: Foster, Timothy J.
; APPLICANT: McDevitt, Damien L.
; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene
; FILE REFERENCE: 05344.105011
; CURRENT APPLICATION NUMBER: US/09/421,868
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/293,728
; FILING DATE: 1994-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-421-868-3
Query Match 59.1%; Score 13; DB 3; Length 12;
Best Local Similarity 40.0%; Pred. No. 4.7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
| |
Db 5 YTSAF 9

RESULT 5
US-07-841-997A-16
; Sequence 16, Application US/07841997A
; Patent No. 5422254
; GENERAL INFORMATION:
; APPLICANT: Lonsborough, John
; APPLICANT: Vuorio, Outi
; TITLE OF INVENTION: A method to increase the trehalose content
; TITLE OF INVENTION: of organisms by transforming them with the
; TITLE OF INVENTION: structural genes for the short and long chains
; TITLE OF INVENTION: yeast trehalose synthase.
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alko Ltd.
; STREET: PO Box 350
; CITY: Helsinki
; STATE: -
; COUNTRY: Finland
; ZIP: SF-00101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WP5.1 file exported as DOS text file
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/07/841,997A  
FILING DATE: 19920228  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/836,021  
FILING DATE: February 14, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary E. Gormley  
REGISTRATION NUMBER: 34409  
REFERENCE/DOCKET NUMBER: 920085A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 659-2930  
TELEFAX: (202) 887-0357  
TELEX: 440142  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-07-841-997A-16

Query Match 59.1%; Score 13; DB 1; Length 14;  
Best Local Similarity 40.0%; Pred. No. 5.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
Db 1 YTSAP 5

## RESULT 6

US-08-290-301-16  
Sequence 16, Application US/08290301  
Patent No. 5792921

## GENERAL INFORMATION:

APPLICANT: Londerborough, John

APPLICANT: Tunnela, Outi

APPLICANT: Palva, Tupio

APPLICANT: Holmstrom, Kjell-Ove

APPLICANT: Welin, Bjorn

APPLICANT: Mandel, Abul

TITLE OF INVENTION: Increasing the trehalose content

TITLE OF INVENTION: of organisms by transforming them with combinations of

TITLE OF INVENTION: the structural genes for trehalose synthase.

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Alko Ltd.

STREET: PO Box 350

CITY: Helsinki

STATE: -

COUNTRY: Finland

ZIP: SF-00101

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb

COMPUTER: IBM PC/XT/AT

OPERATING SYSTEM: PC-DOS

SOFTWARE: WP5.1 file exported as DOS text file

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,301

FILING DATE: 15 August 1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FI 943133

FILING DATE: 29 June 1994

APPLICATION NUMBER: PCT/FI93/00049

FILING DATE: 15 February 1993

APPLICATION NUMBER: 07/841,997

FILING DATE: 28 February 1992

APPLICATION NUMBER: 07/836,021

FILING DATE: 14 February 1992

ATTORNEY/AGENT INFORMATION:  
NAME: Kubovcik, Ronald J.  
NAME: Lydon, James C.  
REGISTRATION NUMBER: 25,401  
REFERENCE/DOCKET NUMBER: 30,082  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 467-6300  
TELEFAX: (202) 466-2006  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-290-301-16

Query Match 59.1%; Score 13; DB 1; Length 14;  
Best Local Similarity 40.0%; Pred. No. 5.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
Db 1 YTSAP 5

## RESULT 7

US-09-013-598-16

Sequence 16, Application US/09013598

Patent No. 6323001

GENERAL INFORMATION:

APPLICANT: Londerborough, John

APPLICANT: Tunnela, Outi

APPLICANT: Palva, Tupio

APPLICANT: Holmstrom, Kjell-Ove

APPLICANT: Welin, Bjorn

APPLICANT: Mandel, Abul

TITLE OF INVENTION: Increasing the trehalose content

TITLE OF INVENTION: of organisms by transforming them with combinations of

TITLE OF INVENTION: the structural genes for trehalose synthase.

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Alko Ltd.

STREET: PO Box 350

CITY: Helsinki

STATE: -

COUNTRY: Finland

ZIP: SF-00101

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb

COMPUTER: IBM PC/XT/AT

OPERATING SYSTEM: PC-DOS

SOFTWARE: WP5.1 file exported as DOS text file

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/013,598

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/290,301

FILING DATE:

APPLICATION NUMBER: PCT/FI93/00049

FILING DATE: 15 February 1993

APPLICATION NUMBER: 07/841,997

FILING DATE: 28 February 1992

APPLICATION NUMBER: 07/836,021

FILING DATE: 14 February 1992

ATTORNEY/AGENT INFORMATION:

NAME: Kubovcik, Ronald J.

NAME: Lydon, James C.

REGISTRATION NUMBER: 25,401

REGISTRATION NUMBER: 30,082

```

; REFERENCE/DOCKET NUMBER: LAIN-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 467-6300
; TELEFAX: (202) 466-2006
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: Amino acid
; TOPOLOGY: Linear
; MOLECULE TYPE: Peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; US-09-013-598-16

```

```

Query Match          59.1%; Score 13; DB 4; Length 14;
Best Local Similarity 40.0%; Pred. No. 5.2e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 5 YXXXF 9
Db 1 YTSAF 5

```

```

RESULT 8
US-08-031-538-43
; Sequence 43, Application US/08031538
; Patent No. 5968817
; GENERAL INFORMATION:
; APPLICANT: Sutcliffe, J Gregor
; APPLICANT: Erlander, Mark G
; APPLICANT: Lovenberg, Timothy W
; TITLE OF INVENTION: HUMAN SEROTONIN RECEPTORS, DNA ENCODING
; TITLE OF INVENTION: THE RECEPTORS, AND USES THEREOF
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 5968817th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/031,538
; FILING DATE: 19930315
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSP5099P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-031-538-43

```

```

Query Match          59.1%; Score 13; DB 2; Length 15;
Best Local Similarity 40.0%; Pred. No. 5.5e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 5 YXXXF 9
Db 3 YSSAF 7

```

```

RESULT 9
US-08-031-538-56
; Sequence 56, Application US/08031538
; Patent No. 5968817
; GENERAL INFORMATION:
; APPLICANT: Sutcliffe, J Gregor
; APPLICANT: Erlander, Mark G
; APPLICANT: Lovenberg, Timothy W
; TITLE OF INVENTION: HUMAN SEROTONIN RECEPTORS, DNA ENCODING
; TITLE OF INVENTION: THE RECEPTORS, AND USES THEREOF
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 5968817th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/031,538
; FILING DATE: 19930315
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSP5099P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-031-538-56

```

```

Query Match          59.1%; Score 13; DB 2; Length 15;
Best Local Similarity 40.0%; Pred. No. 5.5e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 5 YXXXF 9
Db 3 YSSAF 7

```

```

RESULT 10
US-08-401-512-64
; Sequence 64, Application US/08401512
; Patent No. 5599673
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Curran, Mark E.
; APPLICANT: Wang, Qing
; TITLE OF INVENTION: Long QT Syndrome Genes
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3917
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/401,512  
FILING DATE: 09-MAR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Stephen A.  
REGISTRATION NUMBER: 38,609  
REFERENCE/DOCKET NUMBER: 19780-113879  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4848  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-401-512-64

Query Match 59.1%; Score 13; DB 1; Length 18;  
Best Local Similarity 40.0%; Pred. No. 6.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
DB 10 YSAAF 14

RESULT 11  
US-09-226-012-99  
Sequence 99, Application US/09226012  
Patent No. 6207383  
GENERAL INFORMATION:  
APPLICANT: Keating, Mark T.  
APPLICANT: Splawski, Igor  
TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT  
FILE REFERENCE: 2323-136  
CURRENT APPLICATION NUMBER: US/09/226,012  
CURRENT FILING DATE: 1999-01-06  
EARLIER APPLICATION NUMBER: 09/122,847  
EARLIER FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 99  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-226-012-99

Query Match 59.1%; Score 13; DB 3; Length 18;  
Best Local Similarity 40.0%; Pred. No. 6.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
DB 10 YSAAF 14

RESULT 12  
US-09-017-689A-4  
Sequence 4, Application US/09017689A  
Patent No. 6413940  
GENERAL INFORMATION:  
APPLICANT: AVERBACK, Paul  
TITLE OF INVENTION: PHARMACEUTICALLY ACTIVE AGENTS THAT  
IMPEDE THE FORMATION OF AMYLOID BY IMPEDING THE GENESIS OF

DMS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/017,689A  
FILING DATE: 03-Feb-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/038,694  
FILING DATE: 07-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 018792/0125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-017-689A-4  
Query Match 59.1%; Score 13; DB 4; Length 28;  
Best Local Similarity 40.0%; Pred. No. 8.4e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 5 YXXXF 9  
DB 23 YATTF 27  
RESULT 13  
US-08-620-151-73  
Sequence 73, Application US/08620151  
Patent No. 5928955  
GENERAL INFORMATION:  
APPLICANT: Imperiali, Barbara  
APPLICANT: Walkup, Grant K.  
TITLE OF INVENTION: PEPTIDYL FLUORESCENT CHEMOSENSOR FOR  
NUMBER OF SEQUENCES: 136  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BRINKS, HOFER, GILSON & LIONE  
STREET: NBC Tower - Suite 3600, 455 N. Cityfront  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60611-5599  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/620,151  
FILING DATE: 22-MAR-1996

```
; CLASSIFICATION: 422
; ATTORNEY/AGENT INFORMATION:
; NAME: Shannon, Karen L.
; REGISTRATION NUMBER: 36,675
; REFERENCE/DOCKET NUMBER: 8597/6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-321-4200
; TELEFAX: 312-321-4299
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-620-151-73

Query Match      59.1%; Score 13; DB 2; Length 29;
Best Local Similarity 40.0%; Pred. No. 8.6e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      5 YXXXF 9
Db      13 YTTAF 17

RESULT 14
US-09-708-906-7
; Sequence 7, Application US/09708906
; Patent No. 6528620
; GENERAL INFORMATION:
; APPLICANT: Ayer, Donald E.
; APPLICANT: Billin, Andrew N.
; TITLE OF INVENTION: SID-POLYAMIDE FUSIONS: A POTENT METHOD OF REGULATING
; TITLE OF INVENTION: GENE EXPRESSION
; FILE REFERENCE: 1321.2.37
; CURRENT APPLICATION NUMBER: US/09/708,906
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/163,960
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-708-906-7

Query Match      59.1%; Score 13; DB 4; Length 29;
Best Local Similarity 40.0%; Pred. No. 8.6e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      5 YXXXF 9
Db      24 YASSF 28

RESULT 15
US-09-708-906-8
; Sequence 8, Application US/09708906
; Patent No. 6528620
; GENERAL INFORMATION:
; APPLICANT: Ayer, Donald E.
; APPLICANT: Billin, Andrew N.
; TITLE OF INVENTION: SID-POLYAMIDE FUSIONS: A POTENT METHOD OF REGULATING
; TITLE OF INVENTION: GENE EXPRESSION
; FILE REFERENCE: 1321.2.37
; CURRENT APPLICATION NUMBER: US/09/708,906
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/163,960
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-708-906-8

Query Match      59.1%; Score 13; DB 4; Length 29;
Best Local Similarity 40.0%; Pred. No. 8.6e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      5 YXXXF 9
Db      24 YASSF 28

Search completed: October 5, 2004, 16:15:38
Job time : 16.9178 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:12:48 ; Search time 53.0411 Seconds  
(without alignments)  
66.737 Million cell updates/sec

Title: US-09-973-473A-22  
Perfect score: 22  
Sequence: 1 XXXXXFFXX 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	59.1	6	14 US-10-174-209-70	Sequence 70, Appl
2	13	59.1	7	16 US-10-327-598-506	Sequence 506, Appl
3	13	59.1	9	8 US-08-452-843A-1	Sequence 1, Appl
4	13	59.1	9	10 US-08-344-824-17	Sequence 17, Appl
5	13	59.1	9	8 US-09-977-797A-50	Sequence 50, Appl
6	13	59.1	9	12 US-10-149-135-2190	Sequence 2190, Appl
7	13	59.1	9	12 US-10-149-135-2191	Sequence 2191, Appl
8	13	59.1	9	12 US-10-149-135-2192	Sequence 2192, Appl
9	13	59.1	9	12 US-10-149-135-2193	Sequence 2193, Appl
10	13	59.1	9	12 US-10-149-135-2305	Sequence 2305, Appl
11	13	59.1	9	12 US-10-149-135-2306	Sequence 2306, Appl
12	13	59.1	9	12 US-10-149-135-2307	Sequence 2307, Appl
13	13	59.1	9	12 US-10-149-135-2309	Sequence 2309, Appl
14	13	59.1	9	12 US-10-149-135-2310	Sequence 2310, Appl
15	13	59.1	9	12 US-10-149-135-2311	Sequence 2311, Appl

16	13	59.1	9	15 US-10-149-138-4203	Sequence 4203, Ap
17	13	59.1	9	15 US-10-149-138-4204	Sequence 4204, Ap
18	13	59.1	9	15 US-10-149-138-4205	Sequence 4205, Ap
19	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
20	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
21	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
22	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
23	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
24	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
25	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
26	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
27	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
28	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
29	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
30	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
31	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
32	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
33	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
34	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
35	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
36	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
37	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
38	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
39	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
40	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
41	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
42	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
43	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
44	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap
45	13	59.1	9	15 US-10-149-138-4206	Sequence 4206, Ap

ALIGNMENTS

RESULT 1  
US-10-174-209-70  
; Sequence 70, Application US/10174209  
; Publication No. US20030177526A1  
; GENERAL INFORMATION:  
; APPLICANT: Song, Xiaoling  
; APPLICANT: Bariola, Pauline A.  
; APPLICANT: Linderoth, No. US20030177526A1a A.  
; APPLICANT: Fan, Hao  
; APPLICANT: Wei, Zhong-Min  
; TITLE OF INVENTION: RECEPTORS FOR HYPERSENSITIVE RESPONSE ELICITORS AND  
; FILE OF INVENTION: US09-791-551-15  
; FILE REFERENCE: 21829/211  
; CURRENT APPLICATION NUMBER: US/10/174,209  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: 60/335,776  
; PRIOR FILING DATE: 2001-10-31  
; PRIOR APPLICATION NUMBER: 09/810,997  
; PRIOR FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 86  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 70  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: peptide  
US-10-174-209-70

Query Match 59.1%; Score 13; DB 14; Length 6;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXFX 9  
Db 1 YSSAF 5

```

RESULT 2
US-10-327-598-506
; Sequence 506, Application US/10327598
; Publication No. US20040181039A1
; GENERAL INFORMATION:
; APPLICANT: Krah, Eugene
; APPLICANT: Guo, Honliang
; APPLICANT: Aiyyappa, Ashok
; APPLICANT: Lawton, Robert
; TITLE OF INVENTION: Canine Immunoglobulin Variable Domains, Caninized Antibodies, and
; TITLE OF INVENTION: for Making and Using Them
; FILE REFERENCE: 01-799-A
; CURRENT APPLICATION NUMBER: US/10/327,598
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/344,874
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 1139
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 506
; LENGTH: 7
; TYPE: PRT
; ORGANISM: canis familiaris;
US-10-327-598-506

Query Match          59.1%; Score 13; DB 16; Length 7;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 2 YSSSF 6

RESULT 3
US-08-452-843A-1
; Sequence 1, Application US/08452843A
; Publication No. US20020098197A1
; GENERAL INFORMATION:
; APPLICANT: Sette, Alessandro
; APPLICANT: Sidney, John
; TITLE OF INVENTION: HLA Binding Peptides and Their Uses
; FILE REFERENCE: 399632001321
; CURRENT APPLICATION NUMBER: US/08/452,843A
; CURRENT FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: US 08/344,824
; PRIOR FILING DATE: 1994-11-23
; PRIOR APPLICATION NUMBER: US 08/278,634
; PRIOR FILING DATE: 1994-07-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: B35 consensus peptide
US-08-452-843A-1

Query Match          59.1%; Score 13; DB 8; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 5 YAAAF 9

RESULT 4
US-08-344-824-17
; Sequence 17, Application US/08344824
; Publication No. US20030152580A1
; GENERAL INFORMATION:
; APPLICANT: SETTE, Alessandro

```

```

; APPLICANT: SIDNEY, John
; TITLE OF INVENTION: HLA BINDING PEPTIDES AND THEIR USES
; NUMBER OF SEQUENCES: 399
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/344,824
; FILING DATE: 23-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,634
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 14137-80-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-344-824-17

Query Match          59.1%; Score 13; DB 8; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 5 YAAAF 9

RESULT 5
US-09-977-797A-50
; Sequence 50, Application US/09977797A
; Publication No. US20030044772A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Huse, William D.
; APPLICANT: Wu, Herren
; TITLE OF INVENTION: Methods for Identifying Ligand Specific Binding Molecules
; FILE REFERENCE: AME-06805
; CURRENT APPLICATION NUMBER: US/09/977,797A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 09/129,026
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 08/905,825
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 136
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-797A-50

Query Match          59.1%; Score 13; DB 10; Length 9;

```

```
; Best Local Similarity 40.0%; Pred. No. 1.2e+06; Mismatches 3; Indels 0; Gaps 0;
Matches 2; Conservative 0;

QY 5 YXXXP 9
DB 3 YSSTP 7

RESULT 6
US-10-149-135-2190
; Sequence 2190, Application US/10149135
; Publication No. US20040053822A1
; GENERAL INFORMATION:
; APPLICANT: Fikes, John
; APPLICANT: Sette, Alessandro
; APPLICANT: Sidney, John
; APPLICANT: Southwood, Scott
; APPLICANT: Chesnut, Robert
; APPLICANT: Celis, Esteban
; APPLICANT: Keogh, Elissa
; TITLE OF INVENTION: Inducing Cellular Immune Responses to
; FILE REFERENCE: 2060.0130001
; CURRENT APPLICATION NUMBER: US/10/149,135
; CURRENT FILING DATE: 2000-12-11
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 09/458,298
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-06-04
; PRIOR APPLICATION NUMBER: US 08/027,146
; NUMBER OF SEQ ID NOS: 2479
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2190
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificial Peptide
US-10-149-135-2190

Query Match 59.1%; Score 13; DB 12; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXP 9
DB 5 YAAAF 9

RESULT 7
US-10-149-135-2191
; Sequence 2191, Application US/10149135
; Publication No. US20040053822A1
; GENERAL INFORMATION:
; APPLICANT: Fikes, John
; APPLICANT: Sette, Alessandro
; APPLICANT: Sidney, John
; APPLICANT: Southwood, Scott
; APPLICANT: Chesnut, Robert
; APPLICANT: Celis, Esteban
; APPLICANT: Keogh, Elissa
; TITLE OF INVENTION: Inducing Cellular Immune Responses to
; FILE REFERENCE: 2060.0130001
```

```
; CURRENT APPLICATION NUMBER: US/10/149,135
; CURRENT FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: PCT/US00/33545
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 09/458,298
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-06-04
; PRIOR APPLICATION NUMBER: US 08/027,146
; NUMBER OF SEQ ID NOS: 2479
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2191
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificial Peptide
US-10-149-135-2191

Query Match 59.1%; Score 13; DB 12; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXP 9
DB 5 YAAAF 9

RESULT 8
US-10-149-135-2192
; Sequence 2192, Application US/10149135
; Publication No. US20040053822A1
; GENERAL INFORMATION:
; APPLICANT: Fikes, John
; APPLICANT: Sette, Alessandro
; APPLICANT: Sidney, John
; APPLICANT: Southwood, Scott
; APPLICANT: Chesnut, Robert
; APPLICANT: Celis, Esteban
; APPLICANT: Keogh, Elissa
; TITLE OF INVENTION: Inducing Cellular Immune Responses to
; FILE REFERENCE: 2060.0130001
; CURRENT APPLICATION NUMBER: US/10/149,135
; CURRENT FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: PCT/US00/33545
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 09/458,298
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-06-04
; PRIOR APPLICATION NUMBER: US 08/027,146
; NUMBER OF SEQ ID NOS: 2479
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2192
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2192

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 YXXXF 9  
|  
Db 5 YAAAF 9

## RESULT 9

US-10-149-135-2193  
; Sequence 2193, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Cellis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2193  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide

US-10-149-135-2193  
; Sequence 2193, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Cellis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2193  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 YXXXF 9  
|  
Db 5 YAAAF 9

## RESULT 10

US-10-149-135-2305  
; Sequence 2305, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Cellis, Esteban

; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2305  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide

US-10-149-135-2305  
; Sequence 2305, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Cellis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2306  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 YXXXF 9  
|  
Db 5 YAAAF 9

## RESULT 11

US-10-149-135-2306  
; Sequence 2306, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Cellis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2306

; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2306

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
| |  
Db 5 YAAAF 9

## RESULT 12

US-10-149-135-2307  
; Sequence 2307, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2307  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide

US-10-149-135-2309

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
| |  
Db 5 YAAAF 9

## RESULT 13

US-10-149-135-2309  
; Sequence 2309, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2307  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide

US-10-149-135-2307

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
| |  
Db 5 YAAAF 9

## RESULT 14

US-10-149-135-2310  
; Sequence 2310, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146

; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2309  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide

US-10-149-135-2309

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
| |  
Db 5 YAAAF 9

## RESULT 14

US-10-149-135-2310  
; Sequence 2310, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: 2060.0130001  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146

;  
; PRIOR FILING DATE: 1993-03-05  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2310  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2310

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
Db 5 YAAAF 9

RESULT 15  
US-10-149-135-2311  
; Sequence 2311, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Cellis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE REFERENCE: MAGE2/3 Using Peptide and Nucleic Acid Compositions  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; PRIOR FILING DATE: 1993-03-05  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2311  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2311

Query Match 59.1%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
Db 5 YAAAF 9

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:58:01 ; Search time 13.411 Seconds  
(without alignments)  
78.899 Million cell updates/sec

Title: US-09-973-473a-22  
Perfect score: 22  
Sequence: 1 XXXXXFFXX 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	13	59.1	38	2	S22210	photosystem I prot
2	13	59.1	40	2	H95063	hypothetical prote
3	13	59.1	46	2	S17919	photosystem II pro
4	13	59.1	57	2	G97910	hypothetical prote
5	13	59.1	58	2	C82818	hypothetical prote
6	13	59.1	68	2	G82600	hypothetical prote
7	13	59.1	70	2	I77547	hypothetical prote
8	13	59.1	72	2	G89851	hypothetical prote
9	13	59.1	75	2	H97803	proline/betaine tr
10	13	59.1	77	2	E49786	bacteriocin probab
11	13	59.1	79	2	AD1761	B. subtilis CsbA p
12	13	59.1	84	2	JT0457	hypothetical prote
13	13	59.1	87	2	T03993	hypothetical prote
14	13	59.1	90	2	C81917	hypothetical prote
15	13	59.1	90	2	A81186	hypothetical prote
16	13	59.1	93	2	S27028	hypothetical prote
17	13	59.1	101	2	A13460	bacterial protein
18	13	59.1	102	2	H11254	Ig heavy chain v r
19	13	59.1	108	2	S51958	hypothetical prote
20	13	59.1	108	2	H90122	dna directed RNA p
21	13	59.1	108	2	G72635	hypothetical prote
22	13	59.1	112	2	S58139	gene 9 protein - p
23	13	59.1	112	2	S47298	sulleyisin - Strepto
24	13	59.1	115	2	G71058	hypothetical prote
25	13	59.1	115	2	T38415	very hypothetical
26	13	59.1	116	2	S22553	Ig heavy chain v r
27	13	59.1	117	2	F87673	conserved hypothet
28	13	59.1	119	2	H71035	hypothetical prote
29	13	59.1	119	2	H84991	hypothetical prote

30 13 59.1 119 2 B90074 hypothetical prote  
31 13 59.1 119 2 AG3219 hypothetical prote  
32 13 59.1 120 2 B72519 hypothetical prote  
33 13 59.1 121 2 D70767 hypothetical prote  
34 13 59.1 122 2 D89803 conserved hypothet  
35 13 59.1 122 2 A00327 probable exported  
36 13 59.1 128 2 T41487 very hypothetical  
37 13 59.1 129 2 H90324 conserved hypothet  
38 13 59.1 130 2 E90425 conserved hypothet  
39 13 59.1 131 2 H87452 conserved hypothet  
40 13 59.1 132 2 A90731 hypothetical prote  
41 13 59.1 134 2 S49531 anti-Sm antibody V  
42 13 59.1 134 2 T20516 hypothetical prote  
43 13 59.1 134 2 A88691 protein F4iH10.9  
44 13 59.1 134 2 H82969 hypothetical prote  
45 13 59.1 135 2 T01620 hypothetical prote

ALIGNMENTS

RESULT 1

S22210  
photosystem I protein psal - Synechococcus sp.  
N:Alternate names: photosystem I chain VIII  
C:Species: Synechococcus sp.  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Mar-2001  
C:Accession: S22210  
R:Muehlenhoff, U.; Haehnel, W.; Witt, H.T.; Herrmann, R.G.  
submitted to the EMBL Data Library, January 1992  
A:Description: Genes encoding ten subunits of photosystem I from the thermophilic cyanob  
A:Reference number: S18970  
A:Accession: S22210  
A:Molecule type: DNA  
A:Residues: 1-38 <MUE>  
A:Cross-references: EMBL:X63763; NID:G47585; PIDN:CAA45297.1; PID:G47588  
C:Genetics:  
A:Gene: psal  
C:Superfamily: photosystem I protein psal  
C:Keywords: membrane-associated complex; photosynthesis; photosystem I; thylakoid

Query Match 59.1%; Score 13; DB 2; Length 38;  
Best Local Similarity 40.0%; Pred. No. 4.7e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
Db 5 YAAAF 9

RESULT 2

H95063  
hypothetical protein SP0548 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: H95063  
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I  
non, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, P.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: H95063  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-40 <KUR>  
A:Cross-references: GB:AR005672; PIDN:AAK74705.1; PID:G14972023; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0548

Query Match 59.1%; Score 13; DB 2; Length 40;

Best Local Similarity 40.0%; Pred. No. 4.9e+02; Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY 5 YXXXF 9	
Db 29 YSTSF 33	
<p>RESULT 3</p> <p>S17919</p> <p>hypothetical protein psbK - garden pea chloroplast (fragment)</p> <p>C:Species: chloroplast Pisum sativum (garden pea)</p> <p>C&gt;Date: 13-Jan-1995 #sequence_revision 10-Oct-1997 #text_change 26-Aug-1999</p> <p>C:Accession: S17919; S39471</p> <p>R:Nagano, Y.; Matsuno, R.; Sasaki, Y.</p> <p>Curr. Genet. 20, 431-436, 1991</p> <p>A:Title: Sequence and transcriptional analysis of the gene cluster trnQ-zfpA-psaI-ORF231</p> <p>A:Reference number: S17919; MUID:92224289; PMID:1807835</p> <p>A:Accession: S17919</p> <p>A:Molecule type: DNA</p> <p>A:Residues: 1-29 &lt;NAG&gt;</p> <p>A:Cross-references: EMBL:X56315; NID:g12187; PIDN:CAA39753.1; PID:g12188</p> <p>R:Zakharov, S.D.; Ewy, R.G.; Dilley, R.A.</p> <p>FEMS Lett. 336, 95-99, 1993</p> <p>A:Title: Subunit III of the chloroplast ATP-synthase can form a Ca(2+)-binding site on b</p> <p>A:Reference number: S39470; MUID:94085601; PMID:8262226</p> <p>A:Accession: S39471</p> <p>A:Molecule type: protein</p> <p>A:Residues: 25-46 &lt;ZAK&gt;</p> <p>C:Genetics:</p> <p>A:Gene: psbK</p> <p>A:Genome: chloroplast</p> <p>A:Superfamily: photosystem II protein psbK</p> <p>C:Keywords: chloroplast; membrane-associated complex; photosynthesis; photosystem II; th</p>	
<p>Query Match 59.1%; Score 13; DB 2; Length 46;</p> <p>Best Local Similarity 40.0%; Pred. No. 5.4e+02;</p> <p>Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p>	
QY 5 YXXXF 9	
Db 17 YSSSF 21	
<p>RESULT 4</p> <p>G97910</p> <p>hypothetical protein spr0311 [imported] - Streptococcus pneumoniae (strain R6)</p> <p>C:Species: Streptococcus pneumoniae</p> <p>C&gt;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001</p> <p>C:Accession: G97910</p> <p>R:Hoekings, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; E</p> <p>e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M</p> <p>Y, P.; Sun, P.M.; Winkler, M.E.</p> <p>J. Bacteriol. 183, 5709-5717, 2001</p> <p>A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;</p> <p>A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.</p> <p>A:Reference number: A97872; MUID:21429245; PMID:11544234</p> <p>A:Accession: G97910</p> <p>A&gt;Status: preliminary</p> <p>A:Molecule type: DNA</p> <p>A:Residues: 1-57 &lt;KUR&gt;</p> <p>A:Cross-references: GB:AE007317; PIDN:AAK99115.1; PID:g15457866; GSPDB:GN00174</p> <p>C:Genetics:</p> <p>A:Gene: spr0311</p>	
<p>Query Match 59.1%; Score 13; DB 2; Length 57;</p> <p>Best Local Similarity 40.0%; Pred. No. 6.5e+02;</p> <p>Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p>	
QY 5 YXXXF 9	
Db 33 YSSTF 37	

Best Local Similarity 40.0%; Pred. No. 4.9e+02; Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY 5 YXXXF 9	
Db 29 YSTSF 33	
<p>RESULT 3</p> <p>S2818</p> <p>hypothetical protein XF0336 [imported] - Xylella fastidiosa (strain 9a5c)</p> <p>C:Species: Xylella fastidiosa</p> <p>C&gt;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000</p> <p>C:Accession: G82818</p> <p>R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc</p> <p>Nature 406, 151-157, 2000</p> <p>A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.</p> <p>A:Reference number: A82515; MUID:20365717; PMID:10910347</p> <p>A&gt;Note: for a complete list of authors see reference number A59328 below</p> <p>A:Accession: G82818</p> <p>A&gt;Status: preliminary</p> <p>A:Molecule type: DNA</p> <p>A:Residues: 1-58 &lt;SIM&gt;</p> <p>A:Cross-references: GB:AE003886; GB:AE003849; NID:g9105157; PIDN:AAF83146.1; GSPDB:GN0012</p> <p>A:Experimental source: strain 9a5c</p> <p>R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al</p> <p>Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.</p> <p>as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.</p> <p>submitted to GenBank, June 2000</p> <p>A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme</p> <p>J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigre</p> <p>Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.</p> <p>A:Authors: Martins, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;</p> <p>, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;</p> <p>Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak</p> <p>A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir</p> <p>M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z</p> <p>A:Reference number: A59328</p> <p>A:Contents: annotation</p> <p>C:Genetics:</p> <p>A:Gene: XF0336</p>	
<p>Query Match 59.1%; Score 13; DB 2; Length 58;</p> <p>Best Local Similarity 40.0%; Pred. No. 6.5e+02;</p> <p>Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;</p>	
QY 5 YXXXF 9	
Db 37 YSAAP 41	
<p>RESULT 6</p> <p>G82600</p> <p>hypothetical protein XF2098 [imported] - Xylella fastidiosa (strain 9a5c)</p> <p>C:Species: Xylella fastidiosa</p> <p>C&gt;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000</p> <p>C:Accession: G82600</p> <p>R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc</p> <p>Nature 406, 151-157, 2000</p> <p>A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.</p> <p>A:Reference number: A82515; MUID:20365717; PMID:10910347</p> <p>A&gt;Note: for a complete list of authors see reference number A59328 below</p> <p>A:Accession: G82600</p> <p>A&gt;Status: preliminary</p> <p>A:Molecule type: DNA</p> <p>A:Residues: 1-68 &lt;SIM&gt;</p> <p>A:Cross-references: GB:AE004025; GB:AE003849; NID:g9107217; PIDN:AAF84897.1; GSPDB:GN0012</p> <p>A:Experimental source: strain 9a5c</p> <p>R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al</p> <p>Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.</p> <p>as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.</p> <p>submitted to GenBank, June 2000</p> <p>A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme</p> <p>J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigre</p> <p>Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.</p> <p>A:Authors: Martins, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;</p> <p>, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;</p> <p>Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak</p> <p>A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir</p> <p>M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z</p>	



A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF2098

Query Match 59.1%; Score 13; DB 2; Length 68;  
Best Local Similarity 40.0%; Pred. No. 7.5e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
|  
Db 28 YAAFP 32

## RESULT 7

177547  
hypothetical protein 2 - Escherichia coli insertion sequence IS903

C;Species: Escherichia coli  
C;Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 20-Sep-1999  
C;Accession: I77547  
R;Mollet, B.; Iida, S.; Arber, W.  
Mol. Gen. Genet. 199, 534-536, 1985  
A;Title: An active variant of the prokaryotic transposable element IS903 carries an amb  
A;Reference number: I77738; MUID:85295477; PMID:2993802  
A;Accession: I77547  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-70 <RES>  
A;Cross-references: EMBL:X02527; NID:G43694; PIDN:CAA26363.1; PID:G43696  
A;Experimental source: strain K-12, subspecies WA921  
C;Genetics:  
C;Superfamily: hypothetical protein IR903

Query Match 59.1%; Score 13; DB 2; Length 70;  
Best Local Similarity 40.0%; Pred. No. 7.7e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
|  
Db 14 YTTAP 18

## RESULT 8

G89851  
hypothetical protein [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: G89851  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: G89851  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-72 <KUR>  
A;Cross-references: GB:BA000018; PID:gl3700673; PIDN:BA841970.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: SA0737

Query Match 59.1%; Score 13; DB 2; Length 72;  
Best Local Similarity 40.0%; Pred. No. 7.8e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
|  
Db 63 YTTSP 67

## RESULT 9

H97803  
proline/betaine transporter RC0832 homolog [imported] - Rickettsia conorii (strain Maliss  
C;Species: Rickettsia conorii  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
C;Accession: H97803  
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
Science 293, 2093-2098, 2001  
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A;Reference number: A97700; MUID:21442074; PMID:11557893  
A;Accession: H97803  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-75 <KUR>  
A;Cross-references: GB:AR006914; PIDN:AAL03370.1; PID:gl5619933; GSPDB:GN00173  
C;Genetics:  
A;Gene: RC0832

Query Match 59.1%; Score 13; DB 2; Length 75;  
Best Local Similarity 40.0%; Pred. No. 8.1e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
|  
Db 59 YSSAP 63

## RESULT 10

E49786  
bacteriocin probable secretion protein A2 - Lactococcus lactis subsp. cremoris (strain 91  
C;Species: Lactococcus lactis subsp. cremoris  
C;Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 20-Mar-1998  
C;Accession: E49786  
R;van Belkum, M.J.; Hayema, B.J.; Jeeninga, R.E.; Kok, J.; Venema, G.  
Appl. Environ. Microbiol. 57, 492-498, 1991  
A;Title: Organization and nucleotide sequences of two lactococcal bacteriocin operons.  
A;Reference number: A49786; MUID:91197113; PMID:1901707  
A;Accession: E49786  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-77 <VAN>  
C;Genetics:  
A;Genome: plasmid

Query Match 59.1%; Score 13; DB 2; Length 77;  
Best Local Similarity 40.0%; Pred. No. 8.3e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 YXXXF 9  
|  
Db 12 YSSSF 16

## RESULT 11

AD1761  
B. subtilis CsbA protein homolog lin2633 [imported] - Listeria innocua (strain Clp11262)  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AD1761  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fslhi, H.;  
Science 294, 849-852, 2001  
A;Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AD1761  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-79 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC97860.1; PID:gl6415170; GSPDB:GN00178  
A;Experimental source: strain Clp11262

## C:Genetics:

A:Gene: lin2633

Query Match 59.1%; Score 13; DB 2; Length 79;  
 Best Local Similarity 40.0%; Pred. No. 8.5e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 44 YTSSTF 48

## RESULT 12

JT0457

hypothetical protein, 9K - fowlpox virus (isolate HP-438[Munich])

N:Alternate names: hypothetical protein c

C:Species: fowlpox virus

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 21-Jul-2000

C:Accession: JT0457

R:Tomley, F.; Binns, M.; Campbell, J.; Boursnell, M.

J. Gen. Virol. 69, 1025-1040, 1988

A:Title: Sequence analysis of an 11.2 kilobase, near-terminal, BamHI fragment of fowlpox

A:Reference number: JT0442; MUID:88229622; PMID:2836548

A:Accession: JT0457

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-84 &lt;TOM&gt;

A:Cross-references: GB:D00295; NID:g221380; PIDN:BAA00195.1; PID:g221386

Query Match 59.1%; Score 13; DB 2; Length 84;  
 Best Local Similarity 40.0%; Pred. No. 8.9e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 30 YSSTF 34

## RESULT 13

T03993

hypothetical protein T5L19.20 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999

C:Accession: T03993

R:Revan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15184

A:Accession: T03993

A:Molecule type: DNA

A:Residues: 1-87 &lt;BEV&gt;

A:Cross-references: EMBL:AL049481

A:Experimental source: cultivar Columbia; BAC clone T5L19

## C:Genetics:

A:Map position: 4

A:Note: T5L19.20

Query Match 59.1%; Score 13; DB 2; Length 87;  
 Best Local Similarity 40.0%; Pred. No. 9.2e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 11 YSSSF 15

## RESULT 14

C81917

hypothetical protein NMA0737 [imported] - Neisseria meningitidis (strain Z2491 serogroup

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: C81917

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; MUID:2022556; PMID:10761919

A:Accession: C81917

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-90 &lt;PAR&gt;

A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84021.1; PID:g7379459

A:Experimental source: serogroup A, strain Z2491

## C:Genetics:

A:Gene: NMA0737

Query Match 59.1%; Score 13; DB 2; Length 90;  
 Best Local Similarity 40.0%; Pred. No. 9.4e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 10 YAAAF 14

## RESULT 15

A81186

hypothetical protein NMB0555 [imported] - Neisseria meningitidis (strain MC58 serogroup I

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C:Accession: A81186

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: A81186

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-90 &lt;TEST&gt;

A:Cross-references: GB:AE002411; GB:AE002098; NID:g7225776; PIDN:AAF40983.1; PID:g7225781

A:Experimental source: serogroup B, strain MC58

## C:Genetics:

A:Gene: NMB0555

Query Match 59.1%; Score 13; DB 2; Length 90;  
 Best Local Similarity 40.0%; Pred. No. 9.4e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 10 YTAAF 14

Search completed: October 5, 2004, 16:13:53

Job time : 15.411 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:36 ; Search time 7.68493 Seconds  
(without alignments)  
74.532 Million cell updates/sec

Title: US-09-973-473A-22  
Perfect score: 22  
Sequence: 1 XXXXXFFXX 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues 141681  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	59.1	38	1	PSAI_SYNEL
2	13	59.1	38	1	PSAI_SYN2
3	13	59.1	46	1	P8BK_PEA
4	13	59.1	59	1	P8BK_SPIOL
5	13	59.1	89	1	CFA_CITFR
6	13	59.1	93	1	YX1_PAVLU
7	13	59.1	108	1	FLO9_YEAST
8	13	59.1	115	1	YD65_SCHPO
9	13	59.1	119	1	Y512_BUCAP
10	13	59.1	119	1	Y531_BUCAI
11	13	59.1	120	1	USH3_HUMAN
12	13	59.1	121	1	YK87_MYCTU
13	13	59.1	131	1	YLU7_CAELU
14	13	59.1	148	1	CADF_DROME
15	13	59.1	149	1	Y38A_MYCGE
16	13	59.1	154	1	YB17_MYCTU
17	13	59.1	154	1	YL66_ARCFU
18	13	59.1	157	1	Y012_BPL2
19	13	59.1	161	1	Y2CX_ECOLI
20	13	59.1	162	1	KCH2_CAVPO
21	13	59.1	163	1	YCBL_BACUN
22	13	59.1	166	1	ING_CAPHI
23	13	59.1	166	1	ING_SHEEP
24	13	59.1	167	1	VHR2_YABAM
25	13	59.1	175	1	LIGT_ECOLI
26	13	59.1	178	1	DUSP_MYXVL
27	13	59.1	178	1	VHR2_YLDV
28	13	59.1	181	1	RL5_METVA
29	13	59.1	183	1	SPF4_BOVIN
30	13	59.1	193	1	NU2M_PORTE
31	13	59.1	194	1	PUR0_HALMI
32	13	59.1	205	1	HEM2_CLOJO
33	13	59.1	206	1	YC94_MYCPN

#### ALIGNMENTS

##### RESULT 1

```
PSAI_SYNEL STANDARD; PRT; 38 AA.
ID P25900;
AC P25900;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem I reaction center subunit VIII.
GN PSAI OR TSR2405.
OS Synechococcus elongatus (Thermosynechococcus elongatus), and
OS Synechococcus elongatus naegeli.
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046, 1141;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.elongatus; STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Ikeno T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.e.naegeli;
RX MEDLINE=9325282; PubMed=8486290;
RA Muehlenhoff U., Haehnel W., Witt H.T., Herrmann R.G.;
RT "Genes encoding eleven subunits of photosystem I from the
RT thermophilic cyanobacterium Synechococcus sp.";
RL Gene 127:71-78(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (4.0 ANGSTROMS).
RC SPECIES=S.e.naegeli;
RX MEDLINE=97057537; PubMed=8901876;
RA Krauss N., Schubert W.D., Klukas O., Fromme P., Witt H.T., Saenger W.;
RT "Photosystem I at 4-A resolution represents the first structural
RT model of a joint photosynthetic reaction centre and core antenna
RT system.";
RL Nat. Struct. Biol. 3:965-973(1996).
CC -!- FUNCTION: May help in the organization of the psal subunit.
CC -!- SIMILARITY: Belongs to the psal family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AP005377; BAC09957.1; --
DR EMBL; X63763; CA945297.1; --
DR PDB; 2PPS; 27-MAY-98.
DR HAMAP; MF_00431; -; 1.
```

```
34 13 59.1 208 1 XLR_MOUSE
35 13 59.1 216 1 NKGD_HUMAN
36 13 59.1 216 1 NKGD_MACMU
37 13 59.1 219 1 EUTL_ECOLI
38 13 59.1 219 1 EUTL_SALTY
39 13 59.1 223 1 TPIS_AERPE
40 13 59.1 227 1 CUS3_YEAST
41 13 59.1 227 1 RHON_MOUSE
42 13 59.1 227 1 RHON_MOUSE
43 13 59.1 228 1 GLUC_COREF
44 13 59.1 228 1 GLUC_CORGL
45 13 59.1 228 1 MX11_HUMAN
P05531 mus musculu
P26718 homo sapien
Q9mzj7 macaca mula
P76541 escherichia
O9zfu9 salmonella
O9ybr1 aeropyrum p
P47001 saccharomyc
P52198 homo sapien
Q9gym5 mus musculu
Q8rq15 corynebacte
P48244 corynebacte
P50539 homo sapien
```

```

DR InterPro; IPR001302; PSI 8.
DR Pfam; PF00796; PSI 8; 1.
DR ProDom; PD003995; PSI 8; 1.
KW Photosystem I; Photosynthesis; Transmembrane; 3D-structure;
KW Complete proteome. 12 32 POTENTIAL.
FT TRANSMEM 12 32
SQ SEQUENCE 38 AA; 4297 MW; E9B0178560DB5CF5 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 38;
Best Local Similarity 40.0%; Pred. No. 2.8e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
Db 5 YAAAF 9

RESULT 2
PSAI_SYNP2
ID PSAI_SYNP2 STANDARD; PRT; 38 AA.
AC Q54752;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem I reaction center subunit VIII.
GN PSAI.
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32049;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96271772; PubMed=8787020;
RA Schlachter W.M., Shen G., Zhao J., Bryant D.A.;
RT "Characterization of psal and psal mutants of Synechococcus sp.
RT strain PCC 7002: a new model for state transitions in
RT cyanobacteria.";
RL Photochem. Photobiol. 64:53-66(1996).
CC -!- SIMILARITY: Belongs to the psal family.
CC -!- FUNCTION: May help in the organization of the psal subunit.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U58035; AAB18909.1; -.
CC HAMAP; MF 00431; -; 1.
CC InterPro; IPR001302; PSI 8.
CC Pfam; PF00796; PSI 8; 1.
KW Photosystem I; Photosynthesis; Transmembrane.
FT TRANSMEM 9 29
SQ SEQUENCE 38 AA; 3957 MW; BE58C30EFAB31832 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 38;
Best Local Similarity 40.0%; Pred. No. 2.8e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
Db 5 YAAAF 9

RESULT 3
PSBK_PEA
ID PSBK_PEA STANDARD; PRT; 46 AA.
AC P28642; Q972J7;
DT 01-DEC-1992 (Rel. 24, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem II reaction center protein K precursor (PSII-K).
DE PSBK.
OS Spinacia oleracea (Spinach).

```

---

```

DE InterPro; IPR001302; PSI 8.
GN PSBK.
OS Pisum sativum (Garden pea).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustoids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE OF 1-29 FROM N.A.
RC STRAIN=cv. Alaska; PubMed=1807835;
RX MEDLINE=92224289; PubMed=1807835;
RA Nagano Y., Matsuno R., Sasaki Y.;
RT "Sequence and transcriptional analysis of the gene cluster trnQ-zfpA-
RT psal-ORF231-petA in pea chloroplasts.";
RL Curr. Genet. 20:431-436(1991).
RN [2]
RP SEQUENCE OF 25-32, AND MASS SPECTROMETRY.
RX MEDLINE=98298118; PubMed=9632665;
RA Zheleva D., Sharma J., Panico M., Morris H.R., Barber J.;
RT "Isolation and characterization of monomeric and dimeric CP47-reaction
RT center photosystem II complexes.";
RL J. Biol. Chem. 273:16122-16127(1998).
RN [3]
RP SEQUENCE OF 25-46.
RX MEDLINE=94085601; PubMed=8262226;
RA Zakharov S.D., Ewy R.G., Dilley R.A.;
RT "Subunit III of the chloroplast ATP-synthase can form a Ca(2+)-binding
RT site on the luminal side of the thylakoid membrane.";
RL FEBS Lett. 336:95-99(1993).
CC -!- FUNCTION: This protein is a component of the reaction center of
CC photosystem II.
CC -!- SIMILARITY: Belongs to the psbK family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X56315; CAA39753.1; -.
CC PIR; S17919; S17919.
CC HAMAP; MF 00441; -; 1.
CC InterPro; IPR003687; PSII_PsbK.
CC Pfam; PF02533; PsbK; 1.
KW Photosystem II; Chloroplast.
FT PROPEP 1 24
FT CHAIN 25 >46
FT CONFLICT 31 31
FT NON_TER 46 46
SQ SEQUENCE 46 AA; 5126 MW; 382B64B0C5417633 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 46;
Best Local Similarity 40.0%; Pred. No. 3.3e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
Db 17 YSSSF 21

RESULT 4
PSBK_SPIOL
ID PSBK_SPIOL STANDARD; PRT; 59 AA.
AC P12163; Q9M3M8;
DT 01-OCT-1989 (Rel. 12, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem II reaction center protein K precursor (PSII-K).
GN PSBK.
OS Spinacia oleracea (Spinach).

```

Chloroplast.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Caryophyllales; Amaranthaceae; Spinacia.  
 NCBI\_TaxID=3562;  
 [1]  
 SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.  
 Murata N., Miyao M., Hayashida N., Hidaka T., Sugitara M.;  
 "Identification of a new gene in the chloroplast genome encoding a  
 low-molecular-mass polypeptide of photosystem II complex.";  
 FEBS Lett. 235:283-288(1988).  
 [2]  
 SEQUENCE FROM N.A.  
 RC STRAIN=ev. Geant d'hiver, and cv. Monatol;  
 RX MEDLINE=21187424; PubMed=11292076;  
 RA Schmitz-Linneweber C., Maier R.M., Alcaraz J.-P., Cottet A.,  
 RA Herrmann R.G., Mache R.;  
 RT "The plastid chromosome of spinach (Spinacia oleracea): complete  
 RT nucleotide sequence and gene organization.";  
 RL Plant Mol. Biol. 45:307-315(2001).  
 [3]  
 SEQUENCE OF 23-34.  
 RC Schroeder W.P., Henrysson T., Akerlund H.E.;  
 RA "Characterization of low molecular mass proteins of photosystem II by  
 RT N-terminal sequencing.";  
 RL FEBS Lett. 235:289-292(1988).  
 [4]  
 SEQUENCE OF 23-36.  
 RX MEDLINE=89121082; PubMed=2644131;  
 RA Ikeuchi M., Takio K., Inoue Y.;  
 RT "N-terminal sequencing of photosystem II low-molecular-mass proteins.  
 RT 5 and 4.1 kDa components of the O2-evolving core complex from higher  
 RT plants.";  
 RL FEBS Lett. 242:263-269(1989).  
 [5]  
 SEQUENCE OF 23-30, AND MASS SPECTROMETRY.  
 RX MEDLINE=92298118; PubMed=9332665;  
 RA Zheleva D., Sharma J., Panico M., Morris H.R., Barber J.;  
 RA "Isolation and characterization of monomeric and dimeric CP47-reaction  
 RT center photosystem II complexes.";  
 RL J. Biol. Chem. 273:16122-16127(1998).  
 CC -!- FUNCTION: This protein is a component of the reaction center of  
 CC photosystem II.  
 CC -!- MASS SPECTROMETRY: MW=4292.1; METHOD=MALDI; RANGE=23-59.  
 CC -!- SIMILARITY: Belongs to the psbK family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; X12786; CAA31277.2; ALT\_INIT.  
 CC DR EMBL; AJ400848; CAB88708.1; -.  
 CC DR HAMAP; MF\_00441; -; 1.  
 CC DR InterPro; IPR003687; PSII\_PsbK.  
 CC DR Pfam; PF02533; PsbK; 1  
 CC KW Photosystem II; Chloroplast.  
 CC FT PROPEP 1 22  
 CC FT CHAIN 23 59 PHOTOSYSTEM II REACTION CENTER PROTEIN K.  
 CC FT CONFLICT 4 4 I -> T (IN REF. 1).  
 CC FT CONFLICT 8 8 I -> IGI (IN REF. 1).  
 CC FT CONFLICT 13 15 ALY -> TLF (IN REF. 1).  
 CC FT CONFLICT 32 32 S -> N (IN REF. 1).  
 CC FT CONFLICT 37 37 F -> I (IN REF. 1).  
 CC SQ SEQUENCE 59 AA; 6749 MW; 25FCFA8925CE157F CRC64;  
 Query Match 59.1%; Score 13; DB 1; Length 59;  
 Best Local Similarity 40.0%; Pred. No. 4e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 YCXB1\_PAVLU STANDARD; PRT; 93 AA.  
 AC Q01572;  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 10.9 kDa protein in SECA 5' region.  
 OS Pavlova lutherii (Monochrysis lutheri).  
 OG Chloroplast.  
 OC Eukaryota; Haptophyceae; Pavlovales; Pavlova.  
 OX NCBI\_TaxID=2832;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93046814; PubMed=1423730;  
 RA Scaramuzzi C.D., Hiller R.G., Stokes H.W.;

QY 5 YXXXF 9  
 Db 15 YSSSF 19  
 RESULT 5  
 CFA\_CITFR STANDARD; PRT; 89 AA.  
 ID P45509;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Cyclopropane-fatty-acyl-phospholipid synthase (EC 2.1.1.79)  
 GN CFA.  
 OS Cyclopropane fatty acid synthase (CFA synthase) (Fragment).  
 OS Citrobacter freundii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Citrobacter.  
 OX NCBI\_TaxID=546;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 30040;  
 RA Daniel R., Gottschalk G.;  
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: TRANSFERS A METHYLENE GROUP FROM S-ADENOSYL-L-METHIONINE  
 CC TO THE CIS DOUBLE BOND OF AN UNSATURATED FATTY ACID CHAIN  
 CC RESULTING IN THE REPLACEMENT OF THE DOUBLE BOND WITH A METHYLENE  
 CC BRIDGE (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + phospholipid  
 CC olefinic fatty acid = S-adenosyl-L-homocysteine + phospholipid  
 CC cyclopropane fatty acid.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; U09771; AAB48842.1; -.  
 CC DR InterPro; IPR003333; CWAS.  
 CC DR Pfam; PF02353; CWAS; 1.  
 CC KW Transferase; Methyltransferase; Lipid synthesis.  
 CC FT NON TER 1  
 CC SQ SEQUENCE 89 AA; 10609 MW; 7D0DB9F377F72EA CRC64;  
 Query Match 59.1%; Score 13; DB 1; Length 89;  
 Best Local Similarity 40.0%; Pred. No. 5.5e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 YXXXF 9  
 Db 46 YSATF 50  
 RESULT 6  
 YCXB1\_PAVLU STANDARD; PRT; 93 AA.  
 ID Q01572;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 10.9 kDa protein in SECA 5' region.  
 OS Pavlova lutherii (Monochrysis lutheri).  
 OG Chloroplast.  
 OC Eukaryota; Haptophyceae; Pavlovales; Pavlova.  
 OX NCBI\_TaxID=2832;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93046814; PubMed=1423730;  
 RA Scaramuzzi C.D., Hiller R.G., Stokes H.W.;

RT "Identification of a chloroplast-encoded *secA* gene homologue in a  
 RT chromophytic alga: possible role in chloroplast protein  
 RT translocation.";

RL Curr. Genet. 22:421-427(1992).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X65961; CAA46775.1; -.

DR PIR; S27028; S27028.

KW Chloroplast; Hypothetical protein.

SQ SEQUENCE 93 AA; 10860 MW; 3F5D17E168B03A4F CRC64;

Query Match 59.1%; Score 13; DB 1; Length 93;

Best Local Similarity 40.0%; Pred. NO. 5.7e+02;

Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXPF 9

Db 57 YTTSF 61

RESULT 7

FL09\_YEAST

ID FL09\_YEAST STANDARD; PRT; 108 AA.

AC P39711;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Flocculation protein FLO9.

GN FLO9 OR YAL064W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

[1]\_TaxID=4932;

RP SEQUENCE FROM N.A.

RC STRAIN=S288c / AB972;

RX MEDLINE=95249563; PubMed=7731988;

RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,

RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,

RA Storms R.K.;

RT "The nucleotide sequence of chromosome I from *Saccharomyces*

RT *cerevisiae*.";

RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).

CC -!- SIMILARITY: Belongs to the flocculin family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U12980; AAC04970.1; -.

DR PIR; S51958; S51958.

DR GeneOnline; 138402; -.

DR SGD; S0000059; FLO9.

SQ SEQUENCE 108 AA; 12772 MW; 2BF3D67501A7B3D2 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 108;

Best Local Similarity 40.0%; Pred. NO. 6.4e+02;

Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXPF 9

Db 3 YATPF 7

RESULT 9

YS12\_BUCAP

ID YS12\_BUCAP STANDARD; PRT; 119 AA.

RESULT 8

YDGS\_SCHPO

ID YDGS\_SCHPO STANDARD; PRT; 115 AA.

AC Q10493;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Very hypothetical protein C26F1.05 in chromosome I.

GN SPAC26F1.05.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Guillermo R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,

RA Certutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of *Schizosaccharomyces pombe*.";

RL Nature 415:871-880(2002).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Z73100; CAA97362.1; -.

DR PIR; T38415; T38415.

DR GeneDB SPombe; SPAC26F1.05; -.

KW Hypothetical protein.

SQ SEQUENCE 115 AA; 13477 MW; B0C842741F30326D CRC64;

Query Match 59.1%; Score 13; DB 1; Length 115;

Best Local Similarity 40.0%; Pred. NO. 6.7e+02;

Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXPF 9

Db 103 YTTSTF 107

```
AC Q8K945;
DT 28-FEB-2003 (Rel. 41, Created)
DR EMBL; AP001119; BABI3224.1; -
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DR HAMAP; MF 00389; -; 1.
DE Hypothetical protein BUs512.
GN BUs512.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamás I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC -!- SIMILARITY: Belongs to the UPF0116 (dsrF) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE014125; AAM68055.1; -
DR HAMAP; MF 00389; -; 1.
DR InterPro; IPR003787; DsrE.
PFam; PF02635; DsrE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 119 AA; 13610 MW; 6EE924AD44058CC0 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 119;
Best Local Similarity 40.0%; Pred. No. 6.9e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 YXXXP 9
Db 61 YTSSP 65

RESULT 10
Y531_BUCAI
ID Y531_BUCAI STANDARD; PRT; 119 AA.
AC P57597;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein BUs31.
GN BUs31.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]_
RP SEQUENCE FROM N.A.
RX STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- SIMILARITY: Belongs to the UPF0116 (dsrF) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF388366; AAL09581.1; -
DR EMBL; AF388368; AAL09582.1; -
DR Genew; HGNC:12605; USH3A.
MIM; 606397; -

or send an email to license@isb-sib.ch).
CC
DR EMBL; AP001119; BABI3224.1; -
DR HAMAP; MF 00389; -; 1.
DR InterPro; IPR003787; DsrE.
PFam; PF02635; DsrE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1119 AA; 13506 MW; CS95FI215BE10938 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 119;
Best Local Similarity 40.0%; Pred. No. 6.9e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 5 YXXXP 9
Db 61 YTSSP 65

RESULT 11
USH3_HUMAN
ID USH3_HUMAN STANDARD; PRT; 120 AA.
AC P58418;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Usher syndrome type 3 protein.
GN USH3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS USH3 LYS-44 AND
RP 77-ILE-LEU-78 DELINS MET.
RX MEDLINE=21426338; PubMed=11524702;
RA Joensuu T., Haemaelaieinen R., Yuan B., Johnson C., Tegelberg S.,
RA Gasparini P., Zelance L., Pirvola U., Pakarinen L., Lehesjoki A.-E.,
RA de la Chapelle A., Sankila E.-M.;
RT "Mutations in a novel gene with transmembrane domains underlie Usher
RT syndrome type 3.";
RL Am. J. Hum. Genet. 69:673-684(2001).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=A;
CC IsoId=P58418-1; Sequence=Displayed;
CC Name=B;
CC IsoId=P58418-2; Sequence=VSP_004008, VSP_004009;
CC -!- TISSUE SPECIFICITY: Widely expressed. Found in the retina.
CC -!- DISEASE: Defects in USH3A are the cause of Usher syndrome type 3
CC (USH3) [MIM:276902]. Usher syndrome is an autosomal recessive
CC condition, characterized by the association of retinitis
CC pigmentosa with sensorineural deafness. Patients with USH3 have
CC progressive hearing loss, variably present vestibular dysfunction
CC and adult onset retinitis pigmentosa. USH3 is a common form of
CC Usher syndrome in Finland, where it accounts for 42% of all Usher
CC syndrome cases.
CC -!- DATABASE: NAME=Mutations of the USH3A gene;
CC NOTE=Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/ush3mut.htm".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF388366; AAL09581.1; -
DR EMBL; AF388368; AAL09582.1; -
DR Genew; HGNC:12605; USH3A.
MIM; 606397; -
```

```
DR MIM; 276902; -.
KW Transmembrane; Alternative splicing; Vision; Disease mutation;
KW Deafness; Retinitis pigmentosa; Usher syndrome.
FT DOMAIN 1 24 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 25 45 POTENTIAL.
FT DOMAIN 46 58 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 59 79 POTENTIAL.
FT DOMAIN 80 120 CYTOPLASMIC (POTENTIAL).
FT VARSPLIC 9 30 VFPDLKAIKPVSHVNVILFSA -> GYKLCTTGILSLF
FT YGTSNT (in isoform B).
FT VARSPLIC 31 120 /FTID-VSP 004008.
FT VARIANT 44 44 Missing (in isoform B).
FT VARIANT 77 78 /FTID-VAR 012241.
FT VARIANT 78 78 IL -> M (in USH3).
FT VARIANT 78 78 /FTID-VAR 012242.
SQ SEQUENCE 120 AA; 13421 MW; 959B081E7665A2D1 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 120;
Best Local Similarity 40.0%; Pred. No. 7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 109 YTTSF 113

RESULT 12
YK87_MYCTU STANDARD; PRT; 121 AA.
AC Q10696;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein RV2087/MT2148.
GN RV2087 OR MT2148 OR WFCY49.27.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12118036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
CC -1- SIMILARITY: SOME, TO TRANSPOSASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L10986; AAA28017.3; -.
DR PIR; S44805; S44805.
DR WormPep; F10E9.7; CE29494.
DR InterPro; IPR002125; dCMP/cyt deam.
DR Pfam; PF00383; dCMP_cyt_deam; 1.
KW Hypothetical protein.
SQ SEQUENCE 131 AA; 14888 MW; F22AF23E54C6315 CRC64;
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 273966; CAA98199.1; -.
DR EMBL; AB007064; -; NOT_ANNOTATED_CDS.
DR PIR; B70767; B70767.
DR TIGR; MT2148; -.
DR Tuberculist; Rv2087; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 121 AA; 12996 MW; 020E92098EFFC0C4 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 121;
Best Local Similarity 40.0%; Pred. No. 7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 11 YATAF 15

RESULT 13
YL07_CAEEL STANDARD; PRT; 131 AA.
AC P34401;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein F10E9.7 in chromosome III.
GN F10E9.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Haller L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38 (1994).
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L10986; AAA28017.3; -.
DR PIR; S44805; S44805.
DR WormPep; F10E9.7; CE29494.
DR InterPro; IPR002125; dCMP/cyt deam.
DR Pfam; PF00383; dCMP_cyt_deam; 1.
KW Hypothetical protein.
SQ SEQUENCE 131 AA; 14888 MW; F22AF23E54C6315 CRC64;
```



Query Match 59.1%; Score 13; DB 1; Length 131;  
 Best Local Similarity 40.0%; Pred. No. 7.5e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXP 9  
 DB 69 YTSFF 73

RESULT 14  
 ID CADP\_DROME STANDARD; PRT; 148 AA.

AC P45594; Q9W1C4;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cofilin/actin depolymerizing factor homolog (D61 protein) (Twinstar protein).  
 DE TSR OR CADP OR CG4254.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94240181; PubMed=8183953;  
 RA Edwards K.A., Montague R.A., Shepard S., Edgar B.A., Erikson R.L., Kiehart D.P.;  
 RT Identification of Drosophila cytoskeletal proteins by induction of abnormal cell shape in fission yeast.  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4589-4593(1994).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R;  
 RX MEDLINE=96095784; PubMed=8522587;  
 RA Gonsalus K.C., Bonaccorsi S., Williams E., Verni F., Gatti M., Goldberg M.L.;  
 RT "Mutations in twinstar, a Drosophila gene encoding a cofilin/ADF homologue, result in defects in centrosome migration and cytokinesis."  
 RT J. Cell Biol. 131:1243-1259(1995).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Frankkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S., Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J., Hooten D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: Belongs to the actin-binding proteins ADF family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U08217; AAA19856.1; -;  
 DR EMBL; U24490; AAC46962.1; -;  
 DR EMBL; U24676; AAC46963.1; -;  
 DR EMBL; AE003462; AAF47146.1; -;  
 DR PIR; A57569; A57569;  
 DR HSP; Q39250; 1F7S;  
 DR FlyBase; FBgn0011726; tsr;  
 DR GO; GO:0003779; F:actin binding; IMP.  
 DR InterPro; IPR002108; Actbind\_cofln.  
 DR Pfam; PF00241; cofilin\_ADF; 1.  
 DR ProDom; PD002129; Actbind\_cofln; 1.  
 DR SMART; SM00102; ADF; 1.  
 DR PROSITE; PS00325; ACTIN DEPOLYMERIZING; 1.  
 KW Actin-binding; Cytoskeleton; Nuclear protein.  
 FT DOMAIN 19 23 ACTIN LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 96 115 ACTIN-BINDING (POTENTIAL).  
 SQ SEQUENCE 148 AA; 17153 MW; 24F7216033859620 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 148;  
 Best Local Similarity 40.0%; Pred. No. 8.2e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXP 9  
 DB 107 YSSSF 111

RESULT 15  
 Y38A MYCGE STANDARD; PRT; 149 AA.

AC Q92B71;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MG384.1.  
 GN MG384.1.  
 OS Mycoplasma genitalium.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33530 / G-37;  
 RX MEDLINE=96026346; PubMed=7569993;  
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L., Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
 RT "The minimal gene complement of Mycoplasma genitalium."

```

RL Science 270:397-403(1995).
RN [2]
RP IDENTIFICATION
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39720; AAC71616.1; -.
DR TIGR; MG384.1; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
SQ SEQUENCE 149 AA; 17731 MW; F0A7A8DD2562384 CRC64;

Query Match 59.1%; Score 13; DB 1; Length 149;
Best Local Similarity 40.0%; Pred. No. 8.3e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
Db 47 YSTAF 51

```

Search completed: October 5, 2004, 16:07:28  
Job time : 9.68493 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:56:41 ; Search time 42.7945 Seconds  
(without alignments)  
81.102 Million cell updates/sec

Title: US-09-973-473A-22

Perfect score: 22

Sequence: 1 XXXXXXXXXFX 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	59.1	32	Q35494	Q35494 paracentrot
2	13	59.1	39	Q87H11	Q87H11 vibrio para
3	13	59.1	40	Q9JNG3	Q9JNG3 streptococc
4	13	59.1	40	Q97S64	Q97S64 streptococc
5	13	59.1	42	Q8KDF7	Q8KDF7 chlorobium
6	13	59.1	42	Q8FHX5	Q8FHX5 escherichia
7	13	59.1	45	Q9DFK1	Q9DFK1 gillichichys
8	13	59.1	47	Q8KDN9	Q8KDN9 chlorobium
9	13	59.1	49	Q8W6Q6	Q8W6Q6 bacterioph
10	13	59.1	56	Q8EWU8	Q8EWU8 mycoplasma
11	13	59.1	57	Q9U7H8	Q9U7H8 plasmodium
12	13	59.1	57	Q8CZ69	Q8CZ69 streptococc
13	13	59.1	58	Q9PG66	Q9PG66 xylella fas
14	13	59.1	59	Q8GRV3	Q8GRV3 oryza sativ
15	13	59.1	60	Q9DGS1	Q9DGS1 amsacta moo
16	13	59.1	60	Q82PY6	Q82PY6 streptomyc

17	13	59.1	61	5	Q95ZM3	Q95ZM3 caenorhabdi
18	13	59.1	61	10	Q8S8Y4	Q8S8Y4 atropa bell
19	13	59.1	61	12	Q8VIB6	Q8VIB6 hepatitis c
20	13	59.1	68	16	Q9PBP2	Q9PBP2 xylella fas
21	13	59.1	68	16	Q87DA7	Q87DA7 xylella fas
22	13	59.1	69	2	Q9RIH9	Q9RIH9 streptococc
23	13	59.1	69	16	Q8FJK6	Q8FJK6 escherichia
24	13	59.1	69	16	Q827X0	Q827X0 streptomyc
25	13	59.1	70	2	Q48347	Q48347 escherichia
26	13	59.1	72	16	Q99VJ9	Q99VJ9 staphylococ
27	13	59.1	72	16	Q932C6	Q932C6 staphylococ
28	13	59.1	73	5	Q9ND38	Q9ND38 plasmodium
29	13	59.1	73	10	Q9PF49	Q9PF49 oryza sativ
30	13	59.1	74	12	Q9YPP8	Q9YPP8 hepatitis c
31	13	59.1	74	12	Q81696	Q81696 hepatitis c
32	13	59.1	74	12	Q9YPP9	Q9YPP9 hepatitis c
33	13	59.1	74	13	Q8JFV7	Q8JFV7 brachydanio
34	13	59.1	75	9	Q38481	Q38481 bacterioph
35	13	59.1	75	16	Q92HD9	Q92HD9 rickettsia
36	13	59.1	75	16	Q8XHA2	Q8XHA2 clostridium
37	13	59.1	76	12	Q68554	Q68554 hepatitis c
38	13	59.1	79	16	Q928A3	Q928A3 listeria in
39	13	59.1	81	5	Q8T3U0	Q8T3U0 drosophila
40	13	59.1	81	17	Q8TLN7	Q8TLN7 methanosarc
41	13	59.1	81	17	Q8Q669	Q8Q669 methanosarc
42	13	59.1	82	2	Q9X5W4	Q9X5W4 rhodobacter
43	13	59.1	83	16	Q82UE3	Q82UE3 nitrosomona
44	13	59.1	84	12	Q9YFJ9	Q9YFJ9 fowlpox vir
45	13	59.1	84	12	Q80DK4	Q80DK4 hepatitis c

#### ALIGNMENTS

#### RESULT 1

Q35494 ID Q35494 PRELIMINARY; PRT; 32 AA.

AC Q35494;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Cytochrome oxidase subunit 3 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide III) (Fragment).

GN COIII.

OS Paracentrotus lividus (Common sea urchin).

OG Mitochondrion.

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

OC Echinoida; Euechinoidea; Echinacea; Echinoida; Echinidae;

OC Paracentrotus.

OX NCBI\_TaxID=7656;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87248108; PubMed=3596250;

RA Cantatore P., Roberti M., Morisco P., Rainaldi G., Gadaleta M.N.,

RA Saccone C.;

RT "A novel gene order in the Paracentrotus lividus mitochondrial genome.";

RL Gene 53:41-54(1987).

CC -|- FUNCTION: SUBUNITS I, II AND III FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX (BY SIMILARITY).

CC -|- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.

CC -|- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 3 FAMILY.

DR EMBL; M16524; AAA31995.2; -

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005739; C:mitochondrion; IEA.

DR GO; GO:0009481; F:aa3-type cytochrome c oxidase; IEA.

DR GO; GO:0009482; F:aa3-type cytochrome c oxidase; IEA.

DR GO; GO:0009483; F:caa3-type cytochrome c oxidase; IEA.

DR GO; GO:0009485; F:caa3-type cytochrome c oxidase; IEA.

DR GO; GO:0004139; F:cytochrome-c oxidase activity; IEA.

DR GO; GO:0016491; F:oxidoreductase activity; IEA.

DR GO; GO:0006118; P:electron transport; IEA.

```

DR InterPro; IPR000298; CytC_oxdse_III.
DR Pfam; PF00510; COX3; 1.
DR ProDom; PD000382; CytC_oxdse_III; 1.
KW Oxidoreductase; Transmembrane; Mitochondrion.
FT NON_TER 1
FT NON_TER 32
SQ SEQUENCE 32 AA; 3793 MW; 6C298A6B440D6A35 CRC64;

Query Match          59.1%; Score 13; DB 8; Length 32;
Best Local Similarity 40.0%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
   |  |  |
Db 25 YSTTF 29

RESULT 2
Q87H11 PRELIMINARY; PRT; 39 AA.
AC Q87H11;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN VPAl154.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae."
RL Lancet 361:743-749(2003).
DR EMBL; AP005088; BAC62497.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 39 AA; 4817 MW; EA265E7CD2141419 CRC64;

Query Match          59.1%; Score 13; DB 16; Length 39;
Best Local Similarity 40.0%; Pred. No. 2.1e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
   |  |  |
Db 23 YASTF 27

RESULT 3
Q9JNG3 PRELIMINARY; PRT; 40 AA.
AC Q9JNG3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sic1.232.
GN SIC.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS6644;
RX MEDLINE=98060444; PubMed=9399523;
RA Mejia L.M., Stockbauer K.E., Pan X., Cravioto A., Musser J.M.;
RT "Characterization of group A Streptococcus strains recovered from
RT Mexican children with pharyngitis by automated DNA sequencing of

```

```

RT virulence-related genes: unexpectedly large variation in the gene
RT (sic) encoding a complement-inhibiting protein."
RL J. Clin. Microbiol. 35:3220-3224(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS6644;
RX MEDLINE=98169508; PubMed=9501227;
RA Stockbauer K.E., Grigsby D., Pan X., Fu Y.X., Mejia L.M., Cravioto A.,
RA Musser J.M.;
RT "Hypervariability generated by natural selection in an extracellular
RT complement-inhibiting protein of serotype M1 strains of group A
RT Streptococcus."
RL Proc. Natl. Acad. Sci. U.S.A. 95:3128-3133(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS6644;
RX MEDLINE=99238853; PubMed=10221878;
RA Hoe N., Nakashima K., Grigsby D., Pan X., Dou S.J., Naidich S.,
RA Garcia M., Kahn E., Bergmire-Sweat D., Musser J.M.;
RT "Rapid molecular genetic subtyping of serotype M1 group A
RT Streptococcus strains."
RL Emerging Infect. Dis. 5:254-263(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS6644;
RX MEDLINE=99353360; PubMed=10426317;
RA Hoe N.P., Nakashima K., Lukowski S., Grigsby D., Liu M., Kordari P.,
RA Dou S.-J., Pan X., Vuopio-Varkila J., Salmelinna S., McGeer A.,
RA Low D.E., Schwartz B., Schuchat A., Naidich S., De Lorenzo D.,
RA Yu Y.-X., Musser J.M.;
RT "Rapid selection of complement-inhibiting protein variants in group A
RT Streptococcus epidemic waves."
RL Nat. Med. 5:924-929(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS6644;
RX PubMed=11015234;
RA Hoe N.P., Kordari P., Cole R., Liu M., Palzkill T., Huang W.,
RA McLellan D., Adams G.J., Hu M., Vuopio-Varkila J., Cate T.R.,
RA Pichichero M.E., Edwards K.M., Eskola J., Low D.E., Musser J.M.;
RT "Human immune response to streptococcal inhibitor of complement, a
RT serotype M1 group A Streptococcus extracellular protein involved in
RT epidemics."
RL J. Infect. Dis. 182:1425-1436(2000).
DR EMBL; AF232537; AAF65001.1; -.
DR InterPro; IPR005328; Sic.
DR Pfam; PF03482; sic; 1.
DR SEQUENCE 40 AA; 4738 MW; 4C250CA3832D06CD CRC64;

Query Match          59.1%; Score 13; DB 2; Length 40;
Best Local Similarity 40.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
   |  |  |
Db 27 YTTSF 31

RESULT 4
Q97S64 PRELIMINARY; PRT; 40 AA.
ID Q97S64
AC Q97S64;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SP0548.
GN SP0548.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=ATCC BAA-334 / TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Ufferback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of *Streptococcus*  
 RT *pneumoniae*."; ;  
 RL Science 293:498-506(2001).  
 DR EMBL; AE007365; AAK74705.1; -.  
 DR PIR; H95063; H95063.  
 DR TIGR; SP0548; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 40 AA; 4563 MW; 52F75CA2F36FF187 CRC64;

Query Match 59.1%; Score 13; DB 16; Length 40;  
 Best Local Similarity 40.0%; Pred. No. 2.2e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 29 YSTSF 33

RESULT 5  
 Q8KDF7 PRELIMINARY; PRT; 42 AA.

AC Q8KDF7;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein CT1093.  
 GN CT1093.  
 OS Chlorobium tepidum.  
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
 OC Chlorobium.  
 OC NCBI\_TaxID=1097;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 49652 / DSM 12025;  
 RX MEDLINE=22103685; PubMed=12093901;  
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
 RA Dodson R.J., DeBoy R., Gwinn M.L., Nelson W.C., Haft D.H.,  
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,  
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,  
 RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,  
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;  
 RT "The complete genome sequence of *Chlorobium tepidum* TUS, a  
 RT photosynthetic, anaerobic, green-sulfur bacterium."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
 DR EMBL; AE012872; AAM72326.1; -.  
 DR TIGR; CT1093; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 42 AA; 4874 MW; 847FA4B0F406CCF1 CRC64;

Query Match 59.1%; Score 13; DB 16; Length 42;  
 Best Local Similarity 40.0%; Pred. No. 2.2e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 29 YSSSF 33

RESULT 6  
 Q8FHX5 PRELIMINARY; PRT; 42 AA.  
 ID Q8FHX5  
 AC Q8FHX5;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN C1702.  
 OS *Escherichia coli* O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 ON NCBI\_TaxID=217992;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=O6.H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic *Escherichia coli*."; ;  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 DR EMBL; AE016760; AAN80169.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 42 AA; 4739 MW; F0571A4C8EE18B01 CRC64;

Query Match 59.1%; Score 13; DB 16; Length 42;  
 Best Local Similarity 40.0%; Pred. No. 2.2e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 20 YASFP 24

RESULT 7  
 Q9DFK1 PRELIMINARY; PRT; 45 AA.

AC Q9DFK1;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE C-4 sterol methyl oxidase (Fragment).  
 OS *Gillichthys seta* (Shortjaw mudsnaker).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidae;  
 OC Gobiidae; Gillichthys.  
 ON NCBI\_TaxID=79683;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Liver;  
 RX MEDLINE=21117151; PubMed=11172064;  
 RA Gracey A.Y., Troll J.V., Somero G.N.;  
 RT "Hypoxia-induced gene expression profiling in the euryoxic fish  
 RT *Gillichthys mirabilis*."; ;  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).  
 DR EMBL; AF266235; AAG13354.1; -.  
 DR InterPro; IPR001064; Crystallin.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 45 AA; 5435 MW; B375920482F4D2C6 CRC64;

Query Match 59.1%; Score 13; DB 13; Length 45;  
 Best Local Similarity 40.0%; Pred. No. 2.4e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9  
 | |  
 Db 15 YASTF 19

RESULT 8  
 Q8KDN9 PRELIMINARY; PRT; 47 AA.  
 ID Q8KDN9  
 AC Q8KDN9;

```

AC Q8KDN9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein CT1006.
GN CT1006.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Debov R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.B., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vamathavan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR EMBL; A8012865; AAM72241.1; -.
DR TIGR; CT1006; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 47 AA; 4897 MW; 6E38DA41374AC45B8 CRC64;

Query Match 59.1%; Score 13; DB 16; Length 47;
Best Local Similarity 40.0%; Pred. No. 2.4e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 10 YTTAF 14

RESULT 9
Q8W6Q6 PRELIMINARY; PRT; 49 AA.
AC Q8W6Q6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Gp44.
GN 44.
OS Bacteriophage phiE125.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=180504;
RN [1]
RP SEQUENCE FROM N.A.
RA Woods D.E., Jeddeloh J.A., Fritz D.F., Deshazer D.;
RT "Burkholderia thailandensis E125 Harbors a Temperate Bacteriophage
RT Specific for Burkholderia mallei.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF447491; AAL40318.1; -.
SQ SEQUENCE 49 AA; 5577 MW; 98F747E0A1D2AC53 CRC64;

Query Match 59.1%; Score 13; DB 9; Length 49;
Best Local Similarity 40.0%; Pred. No. 2.5e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 33 YAAAF 37

RESULT 10
Q8EWU8 PRELIMINARY; PRT; 56 AA.
ID Q8EWU8
AC Q8EWU8;

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN MPE1035.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AF004170; BAC43895.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 56 AA; 6551 MW; EB575D5D435A7A55 CRC64;

Query Match 59.1%; Score 13; DB 16; Length 56;
Best Local Similarity 40.0%; Pred. No. 2.8e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 19 YSTSF 23

RESULT 11
Q9U7H8 PRELIMINARY; PRT; 57 AA.
ID Q9U7H8
AC Q9U7H8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Erythrocyte membrane protein 1 SDI06G (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99404835; PubMed=10477185;
RA Ward C.P., Clotey G.T., Dorris M., Ji D.D., Arnot D.E.;
RT "Analysis of Plasmodium falciparum PfEMP-1/var genes suggests that
RT recombination rearranges constrained sequences.";
RL Mol. Biochem. Parasitol. 102:167-177(1999).
DR EMBL; AF127309; AAD52789.1; -.
FT NON_TER 1
FT NON_TER 1
FT NON_TER 57
SQ SEQUENCE 57 AA; 6304 MW; CD2891670148C8B6 CRC64;

Query Match 59.1%; Score 13; DB 5; Length 57;
Best Local Similarity 40.0%; Pred. No. 2.8e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 YXXXF 9
DB 44 YASSF 48

RESULT 12
Q8CZ69 PRELIMINARY; PRT; 57 AA.
ID Q8CZ69
AC Q8CZ69;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SPRO311.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).

```

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 EX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
 RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
 RA McArthur S.M., McHenry M., Mcleaster K., Mundy C.W., Niclas T.I.,  
 RA Morris P.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,  
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,  
 RA Glaes J.I.;  
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
 RL J. Bacteriol. 183:5709-5717(2001).  
 DR EMBL: AE008412; AAK99115.1; --  
 DR PIR: G97910; G97910.  
 DR InterPro: IPR000437; Prok\_lipoprot.S.  
 DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 57 AA; 6116 MW; 9FDB360477D034E5 CRC64;  
 Query Match 59.1%; Score 13; DB 16; Length 57;  
 Best Local Similarity 40.0%; Pred. No. 2.8e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 YXXXF 9  
 DB 33 YSSTF 37  
 RESULT 13  
 Q9PGG6 PRELIMINARY; PRT; 58 AA.  
 AC Q9PGG6;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein Xf0336.  
 GN XF0336.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAINS9ASC;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Negai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Sanceli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva P.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
 RL Nature 406:151-159(2000).  
 DR EMBL: AE003886; AAF83146.1; --  
 DR PIR: C82818; C82818.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 58 AA; 6888 MW; 4D0944DC6B5D2AA4 CRC64;  
 Query Match 59.1%; Score 13; DB 16; Length 58;  
 Best Local Similarity 40.0%; Pred. No. 2.9e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 YXXXF 9  
 DB 37 YSAAP 41  
 RESULT 14  
 Q8GRV3 PRELIMINARY; PRT; 59 AA.  
 AC Q8GRV3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Glutathione reductase (Fragment).  
 GN RGRC2.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22286393; PubMed=12399401;  
 RA Olsen K.M., Furugganan M.D.;  
 RT "Molecular Evidence on the Origin and Evolution of Glutinous Rice.";  
 RL Genetics 162:941-950(2002).  
 DR EMBL: AY136760; AAN15933.1; --  
 DR EMBL: AY136761; AAN15934.1; --  
 DR EMBL: AY136762; AAN15935.1; --  
 DR EMBL: AY136763; AAN15936.1; --  
 DR EMBL: AY136764; AAN15937.1; --  
 DR EMBL: AY136765; AAN15938.1; --  
 DR EMBL: AY136766; AAN15939.1; --  
 DR GO: GO:0005737; C:cytoplasm; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR Pfam: PF02852; pyr\_redox\_dim.  
 FT NON\_TER 1 59  
 FT TER 59  
 SQ SEQUENCE 59 AA; 6374 MW; B416BD7DB66B391D CRC64;  
 Query Match 59.1%; Score 13; DB 10; Length 59;  
 Best Local Similarity 40.0%; Pred. No. 2.9e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 YXXXF 9  
 DB 47 YTSF 51  
 RESULT 15  
 Q9DGS1 PRELIMINARY; PRT; 60 AA.  
 AC Q9DGS1;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE AMVITR02.  
 GN AMVITR02.

OS Amsacta moorei entomopoxvirus (AmePV).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;  
OC Entomopoxvirus B.  
OX NCBI\_TaxID=28321;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20396580; PubMed=10936094;  
RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,  
RA Moyer R.W.;  
RT "Complete Genomic Sequence of the Amsacta moorei Entomopoxvirus:  
RT Analysis and Comparison with Other Poxviruses."  
RL Virology 274:120-139(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,  
RA Moyer R.W.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF250284; AAG02986.1; -.  
DR EMBL; AF250284; AAG02974.1; -.  
SQ SEQUENCE 60 AA; 6944 MW; FFBD154D5B93AC36 CRC64;  
  
Query Match 59.1%; Score 13; DB 12; Length 60;  
Best Local Similarity 40.0%; Pred. No. 3e+03;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 5 YXXXF 9  
Db 14 YSSSF 18

Search completed: October 5, 2004, 16:12:24  
Job time : 47.7945 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:06 ; Search time 82.1918 Seconds  
(without alignments)  
55.003 Million cell updates/sec

Title: US-09-973-473A-23

Perfect score: 35  
Sequence: 1 XXYYXXXXXXRSP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	60.0	1126	4	ABB69720 Drosophil
2	20	57.1	145	5	ABP30547 Streptoco
3	20	57.1	148	5	ABP27070 Streptoco
4	20	57.1	600	4	AAU14758 Novel bon
5	20	57.1	721	6	ABP71509 Amino aci
6	19	54.3	149	3	AB63124 Human sec
7	19	54.3	191	5	AB94019 Human sec
8	19	54.3	191	5	ABG65126 Human alb
9	19	54.3	193	5	AB94043 Human sec
10	19	54.3	193	5	ABG65124 Human alb
11	19	54.3	351	3	AB63092 Human sec
12	19	54.3	439	4	AG73634 Human col
13	19	54.3	461	5	AB94042 Human sec
14	19	54.3	461	5	ABG65125 Human alb
15	19	54.3	486	4	AB93892 Human pro
16	19	54.3	486	5	ABP68996 Human pol
17	19	54.3	486	5	AAE21058 Human dru
18	19	54.3	486	5	ABU10993 cDNA enco
19	19	54.3	486	6	ABR44422 Human sul
20	19	54.3	486	7	AD34344 Beta-gala
21	19	54.3	524	5	AB94068 Human sec
22	19	54.3	728	6	ABU22550 Protein e
23	19	54.3	1323	7	ADD25207 Fertility
24	19	54.3	1991	4	AB860651 Drosophil
25	19	54.3	2209	1	AAP20037 Sequence

26	18	51.4	70	2	AA68589 Human pro
27	18	51.4	77	4	AAU55658 Propionib
28	18	51.4	77	6	ABMS2177 Propionib
29	18	51.4	94	6	ABM65610 Propionib
30	18	51.4	95	5	ABB06839 Human ncp
31	18	51.4	111	2	AA68588 Human pro
32	18	51.4	149	2	AAV74068 Human pro
33	18	51.4	153	4	AAU54093 Propionib
34	18	51.4	153	6	ABMS0612 Propionib
35	18	51.4	185	3	AAAB11542 SEN virus
36	18	51.4	217	2	AA68588 GGF segme
37	18	51.4	217	2	AA68588 GGF segme
38	18	51.4	217	2	AA68588 GGF segme
39	18	51.4	217	2	AA68588 GGF segme
40	18	51.4	217	2	AA68588 GGF segme
41	18	51.4	217	2	AA68588 GGF segme
42	18	51.4	217	2	AA68588 GGF segme
43	18	51.4	217	2	AA68588 GGF segme
44	18	51.4	217	2	AA68588 GGF segme
45	18	51.4	217	2	AA68588 GGF segme

ALIGNMENTS

RESULT 1  
ABB69720  
ID ABB69720 standard; protein; 1126 AA.  
XX  
AC ABB69720;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 35952.  
XX  
KW Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PI (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-658860/75.  
XX  
DR N-PSDB; ABL13823.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.  
XX  
PS Disclosure; SEQ ID NO 35952; 2lpp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA sequences (AB101840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

```
SQ Sequence 1126 AA;
Query Match 60.0%; Score 21; DB 4; Length 1126;
Best Local Similarity 30.8%; Pred. No. 3e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXRXSP 16
DB 251 YSSTTEESRSRSP 263

RESULT 2
ABP30547
ID ABP30547 standard; protein; 145 AA.
XX AC ABP30547;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 10270.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI: 2002-352536/38.
XX DR N-PSDB; ABN71178.
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX PS Claim 1; Page 4158; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins
XX SQ Sequence 145 AA;
Query Match 57.1%; Score 20; DB 5; Length 145;
Best Local Similarity 30.8%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

SQ Sequence 1126 AA;
Query Match 60.0%; Score 21; DB 4; Length 1126;
Best Local Similarity 30.8%; Pred. No. 3e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXRXSP 16
DB 251 YSSTTEESRSRSP 263

RESULT 3
ABP27070
ID ABP27070 standard; protein; 148 AA.
XX AC ABP27070;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 3316.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI: 2002-352536/38.
XX DR N-PSDB; ABN67701.
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX PS Claim 1; Page 3485; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins
XX SQ Sequence 148 AA;
Query Match 57.1%; Score 20; DB 5; Length 148;
Best Local Similarity 30.8%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

Qy 4 YXXXXXXXXXSP 16  
Db 65 YSLSTATSNRFSF 77

RESULT 4  
AAU14758  
ID AAU14758 standard; protein; 600 AA.  
XX AC AAU14758;  
XX DT 24-OCT-2001 (first entry)  
XX DE Novel bone marrow polypeptide #157.  
XX KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;  
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;  
KW wound healing; nutritional supplement; immune disorder;  
KW severe combined immunodeficiency; SCID.  
XX OS Homo sapiens.  
XX PN WO200157187-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US003782.  
XX PR 03-FEB-2000; 2000US-00496914.  
XX PR 20-JUN-2000; 2000US-00598075.  
XX PR 19-JUL-2000; 2000US-00620325.  
XX PR 30-NOV-2000; 2000US-0250683P.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;  
PI Ren F, Drmanac RT;  
XX WPI; 2001-488875/53.  
XX N-PSDB; AAS23063.  
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and  
XX gene therapy.  
XX Claim 10; Page 136; 392pp; English.

AAU14602-AAU14794 represent novel bone marrow polypeptides of the  
invention. The proteins and corresponding coding sequences may be used in  
the prevention, diagnosis and treatment of diseases associated with  
inappropriate bone marrow polypeptide expression. For example, to treat  
disorders associated with decreased expression by rectifying mutations or  
deletions in a patient's genome that affect the activity of the  
polypeptides by expressing inactive proteins or to supplement the  
patient's own production of the polypeptide. Additionally, the nucleic  
acids may be used to produce the polypeptides, by inserting the nucleic  
acids into a host cell and culturing the cell to express the protein. The  
nucleic acid and its complementary sequences may also be used as DNA  
probes in diagnostic assays to detect and quantitate the presence of  
similar nucleic acid sequences in samples, and therefore which patients  
may be in need of restorative therapy. The proteins may also be used as  
antigens in the production of antibodies against bone marrow proteins and  
in assays to identify modulators of their expression and activity. The  
anti-bone marrow protein antibodies and antagonists may also be used to  
down regulate expression and activity. The antibodies may also be used as  
diagnostic agents for detecting the presence of the protein in samples  
(e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be  
used to regulate haematopoiesis activity, and consequently in the  
treatment of myeloid or lymph cell disorders; in tissue regeneration,  
such as wound healing; as a nutritional supplement; and in treatment of  
immune disorders such as severe combined immunodeficiency (SCID)

Sequence 600 AA;

Query Match 57.1%; Score 20; DB 4; Length 600;  
Best Local Similarity 30.8%; Pred. No. 3.2e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXXXSP 16  
Db 228 YYSRRSSARSSP 240

RESULT 5  
ABP71509  
ID ABP71509 standard; protein; 721 AA.  
XX AC ABP71509;  
XX DT 15-MAY-2003 (first entry)  
XX DE Amino acid sequence of protein AAD41257.  
XX KW Cardiant; hypotension; antiarrhythmic; gene therapy; heart disease;  
KW transgenic; human.  
XX OS Homo sapiens.  
XX PN WO2003006687-A2.  
XX PD 23-JAN-2003.  
XX PF 10-JUL-2002; 2002WO-EP007704.  
XX PR 10-JUL-2001; 2001US-0304385P.  
XX PA (MEDI-) MEDIGENE AG.  
XX PI Reuner B, Bunk D, Henkel T;  
XX WPI; 2003-229493/22.  
XX N-PSDB; ABZ75910.  
XX Identifying a subject at risk for a disease of the heart, comprises  
XX quantitating the amount of at least one RNA or a polypeptide in the heart  
XX tissue or serum of the blood of the subject.  
XX Claim 1; Fig 8C; 197pp; English.

The invention relates to identifying a subject at risk for a disease of  
the heart and involves quantitating the amount of at least one RNA or a  
polypeptide in the heart tissue or serum of the blood of the subject. The  
DNA, polypeptides, compounds identified by the methods above, the refined  
or modified compounds, and the monoclonal antibodies are useful for  
manufacturing a pharmaceutical composition for preventing or treating  
heart diseases, e.g. congestive heart failure, dilative cardiomyopathy,  
hypertrophic cardiomyopathy, ischaemic cardiomyopathy, specific heart  
muscle disease, rhythm and conduction disorders, syncope and sudden  
death, coronary heart disease, systemic arterial hypertension, pulmonary  
congenital heart disease, pericardial disease or endocarditis. Transgenic  
animals are useful for developing medicaments for treating heart  
diseases. The methods are useful for identifying a subject at risk for a  
heart disease, or for identifying compounds for treating heart disease.  
Sequences ABP71501-510 represent specific examples of polypeptides that  
can be quantitated using the method of the invention

Sequence 721 AA;

Query Match 57.1%; Score 20; DB 6; Length 721;  
Best Local Similarity 30.8%; Pred. No. 3.9e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXXXSP 16  
Db 225 YYSRRSSARSSP 237

```
RESULT 6
ID AAB63124 standard; protein; 149 AA.
XX
AC AAB63124;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 44 SEQ ID NO:134.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic;
KW cerebroprotective; neutrotropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnerary; gene therapy; neoplasm;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angogenesis; nervous system disorder; infection;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW skin aging; food additive; preservative.
XX
OS Homo sapiens.
XX
PN WO200061748-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US008982.
XX
PR 09-APR-1999; 99US-0128696P.
XX
PR 14-JAN-2000; 2000US-0176069P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
WPI: 2000-638566/61.
XX
DR New nucleic acid molecules encoding 48 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX PT used as food additives or preservatives.
XX
PS Disclosure; Page 471-472; 480pp; English.
XX
AAAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
XX to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
XX and polypeptides homologous to them. Human secreted proteins have
XX activities based on the tissues and cells the genes are expressed in.
XX Examples of activities include: immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic;
XX cerebroprotective; neutrotropic; neuroprotective; antibacterial; virucide;
XX fungicide; ophthalmological; and vulnerary. The polynucleotides and
XX proteins can be used to prevent, treat or ameliorate a medical condition
XX in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
XX sheep. They are also used in diagnosing a pathological condition or
XX susceptibility to a pathological condition. Disorders which are diagnosed
XX or treated include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischaemia, angogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi and
XX ocular disorders e.g. corneal infection. The polypeptides can also be
XX used to aid wound healing and epithelial cell proliferation, to prevent
XX skin aging due to sunburn, to maintain organs before transplantation, for
XX supporting cell culture of primary tissues, to regenerate tissues and in
XX chemotaxis. The polypeptides can also be used as a food additive or
XX preservative to increase or decrease storage capabilities, fat content,
XX lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
XX nutritional components. AAF22307 to AAF22315 and AAB63048 represent
XX sequences used in the exemplification of the present invention
XX
XX Sequence 149 AA;
XX
Query Match 54.3%; Score 19; DB 5; Length 191;
Best Local Similarity 30.8%; Pred. No. 2.1e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

```
Query Match 54.3%; Score 19; DB 3; Length 149;
Best Local Similarity 30.8%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
DB 105 YYKSTSAFRKSP 117

RESULT 7
AAB94019
ID ABB94019 standard; protein; 191 AA.
XX
AC ABB94019;
XX
DT 06-JUN-2002 (first entry)
XX
DE Human secreted protein SEQ ID NO: 62.
XX
KW Cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200216389-A1.
XX
PD 28-FEB-2002.
XX
PF 17-JAN-2001; 2001WO-US001397.
XX
PR 18-AUG-2000; 2000US-0226281P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Komatsoulis CA, Baker KP, Birse CE, Soppet DR;
XX Olsen HS, Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH;
XX Fiscella M, Ni J;
XX WPI: 2002-292054/33.
XX N-PSDB; ABL92347.
XX
DR Isolated nucleic acids encoding 23 secreted proteins useful for the
XX diagnosis and treatment of e.g. cancer, HIV infection, stroke and
XX rheumatoid arthritis.
XX
PS Claim 11; Page 459-460; 517pp; English.
XX
Sequences ABB94017-94095 represent the amino acid sequences of 79 human
XX secreted proteins encoded by the genes ABL92338-92393. The genes and
XX proteins are useful for preventing, ameliorating or treating medical
XX conditions, e.g. by protein or gene therapy. The genes are isolated from
XX a range of human tissues disclosed in the specification. The nucleic
XX acids, proteins, antibodies and (ant)agonists are useful in the
XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)
XX immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune chryoiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemia; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections
XX
XX Sequence 191 AA;
XX
Query Match 54.3%; Score 19; DB 5; Length 191;
Best Local Similarity 30.8%; Pred. No. 2.1e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

QY 4 YXXXXXXXXXSP 16  
 DB 171 YKSTSTFRKSP 183

## RESULT 8

ABG65126  
 ID ABG65126 standard; protein; 191 AA.

XX AC ABG65126;  
 XX DT 27-AUG-2002 (first entry)  
 XX XX Human albumin fusion protein #1801.

XX KW Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytostatic; antiinfertility; antiinflammatory; antiulcer;  
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.

XX OS Homo sapiens.  
 OS Synthetic.  
 XX PN WO200177137-A1.  
 XX PD 18-OCT-2001.

XX PF 12-APR-2001; 2001WO-US011988.  
 XX PR 12-APR-2000; 2000US-0229358P.  
 XX PR 25-APR-2000; 2000US-0199384P.  
 XX PR 21-DEC-2000; 2000US-0256931P.  
 XX XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Haseltine WA;  
 XX WPI; 2002-010886/01.  
 XX DR New fusion protein for treating disease e.g. diabetes comprises an  
 XX PT albumin fused to a therapeutic protein.

XX PS Claim 1; Page 1772; 2102pp; English.  
 XX CC The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA, also known as human serum  
 CC albumin, HSA). The proteins are useful for treating a disease or disorder  
 CC that may be modulated by therapeutic protein X. The albumin extends the  
 CC shelf-life of protein X, and may increase its biological in vitro/in vivo  
 CC activity. The protein is useful for treating and diagnosing disorders  
 CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
 CC disease, ulcerative colitis), immune disorders (e.g. acquired  
 CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
 CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
 CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
 CC schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
 XX CC ABG63326-ABG65518 represent albumin fusion proteins of the invention

SQ Sequence 191 AA;

Query Match 54.3%; Score 19; DB 5; Length 191;  
 Best Local Similarity 30.8%; Pred. No. 2.1e+03;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXSP 16  
 DB 171 YKSTSTFRKSP 183

## RESULT 9

ABB94043  
 ID ABB94043 standard; protein; 193 AA.

XX AC ABB94043;  
 XX DT 06-JUN-2002 (first entry)  
 XX XX Human secreted protein SEQ ID NO: 86.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW anti-allergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.

XX OS Homo sapiens.  
 XX PN WO200216389-A1.  
 XX PD 28-FEB-2002.

XX PF 17-JAN-2001; 2001WO-US001397.  
 XX PR 18-AUG-2000; 2000US-0226281P.  
 XX XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR;  
 PI Olsen HS, Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH;  
 PI Fiscella M, Ni J;  
 XX WPI; 2002-292054/33.  
 XX DR N-PSDB; ABL92371.

XX PT Isolated nucleic acids encoding 23 secreted proteins useful for the  
 XX PT diagnosis and treatment of e.g. cancer, HIV infection, stroke and  
 XX PT rheumatoid arthritis.  
 XX PS Claim 11; Page 476; 517pp; English.

XX CC Sequences ABB94017-94095 represent the amino acid sequences of 79 human  
 CC secreted proteins encoded by the genes ABL92336-92393. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
 CC a range of human tissues disclosed in the specification. The nucleic  
 CC acids, proteins, antibodies and (ant)agonists are useful in the  
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)  
 CC immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 XX infections

SQ Sequence 193 AA;

Query Match 54.3%; Score 19; DB 5; Length 193;  
 Best Local Similarity 30.8%; Pred. No. 2.1e+03;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXSP 16  
 DB 171 YKSTSTFRKSP 183

## RESULT 10

ABG65124  
 ID ABG65124 standard; protein; 193 AA.

XX

AC ABG65124;  
 XX 27-AUG-2002 (first entry)  
 XX Human albumin fusion protein #1799.  
 DE  
 XX Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytostatic; antiinfertility; antiinflammatory; antiulcer;  
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX WO200177137-A1.  
 XX 18-OCT-2001.  
 PD  
 XX 12-APR-2001; 2001WO-US011988.  
 XX 12-APR-2000; 2000US-0229358P.  
 PR 25-APR-2000; 2000US-0199384P.  
 PR 21-DEC-2000; 2000US-0256931P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Rosen CA, Haseltine WA;  
 PI WPI; 2002-010886/01.  
 DR  
 XX New fusion protein for treating disease e.g. diabetes comprises an  
 PT albumin fused to a therapeutic protein.  
 PT  
 PS Claim 1; Page 1769; 2102pp; English.  
 XX The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA, also known as human serum  
 CC albumin, HSA). The proteins are useful for treating a disease or disorder  
 CC that may be modulated by therapeutic protein X. The albumin extends the  
 CC shelf-life of protein X, and may increase its biological in vitro/in vivo  
 CC activity. The protein is useful for treating and diagnosing disorders  
 CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
 CC disease, ulcerative colitis), immune disorders (e.g. acquired  
 CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
 CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
 CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
 CC schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
 CC ABG63326-ABG65518 represent albumin fusion proteins of the invention  
 XX  
 SQ Sequence 193 AA;  
 Query Match 54.3%; Score 19; DB 5; Length 193;  
 Best Local Similarity 30.8%; Pred. No. 2.1e+03;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 4 YXXXXXXXRXSP 16  
 | | | |  
 Db 171 YKSTSSAFKSP 183  
 RESULT 11  
 AAB63092  
 ID AAB63092 standard; protein; 351 AA.  
 XX  
 AC AAB63092;  
 XX  
 XX 26-MAR-2001 (first entry)  
 DT Human secreted protein sequence encoded by gene 44 SEQ ID NO:102.  
 DE

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
 KW fungicide; ophthalmological; vulnery; gene therapy; neoplasm;  
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
 KW cerebral ischaemia; angiogenesis; nervous system disorder; infection;  
 KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;  
 XX skin aging; food additive; preservative.  
 XX  
 OS Homo sapiens.  
 OS WO2000061748-A1.  
 PN 19-OCT-2000.  
 PD  
 XX 06-APR-2000; 2000WO-US008982.  
 XX 09-APR-1999; 99US-0128696P.  
 PR 14-JAN-2000; 2000US-0176069P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Rosen CA, Ruben SM, Komatsoulis G;  
 PI WPI; 2000-638566/61.  
 DR N-PSDB; AAF22359.  
 DR  
 XX New nucleic acid molecules encoding 48 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 PT  
 PS Claim 11; Page 446-447; 480pp; English.  
 XX  
 XX AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049  
 CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins  
 CC and polypeptides homologous to them. Human secreted proteins have  
 CC activities based on the tissues and cells the genes are expressed in.  
 CC Examples of activities include: immunosuppressive; antiarthritic;  
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 CC cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
 CC fungicide; ophthalmological; and vulnery. The polynucleotides and  
 CC proteins can be used to prevent, treat or ameliorate a medical condition  
 CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or  
 CC sheep. They are also used in diagnosing a pathological condition or  
 CC susceptibility to a pathological condition. Disorders which are diagnosed  
 CC or treated include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and  
 CC ocular disorders e.g. corneal infection. The polypeptides can also be  
 CC used to aid wound healing and epithelial cell proliferation, to prevent  
 CC skin aging due to sunburn, to maintain organs before transplantation, for  
 CC supporting cell culture of primary tissues, to regenerate tissues and in  
 CC chemotaxis. The polypeptides can also be used as a food additive or  
 CC preservative to increase or decrease storage capabilities, fat content,  
 CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other  
 CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent  
 CC sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 351 AA;  
 Query Match 54.3%; Score 19; DB 3; Length 351;  
 Best Local Similarity 30.8%; Pred. No. 3.8e+03;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 4 YXXXXXXXRXSP 16  
 | | | |  
 Db 35 YKSTSSAFKSP 47

RESULT 12  
AAG73634  
ID AAG73634 standard; protein; 439 AA.  
XX AC AAG73634;  
XX DT 03-SEP-2001 (first entry)  
XX DE Human colon cancer antigen protein SEQ ID NO:4398.  
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX KW colorectal carcinoma.  
XX OS Homo sapiens.  
XX PN WO200122920-A2.  
XX PD 05-APR-2001.  
XX PF 28-SEP-2000; 2000WO-US026524.  
XX PR 29-SEP-1999; 99US-0157137P.  
XX PR 03-NOV-1999; 99US-0163280P.  
XX FA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX DR WPI; 2001-235357/24.  
XX DR N-PSDB; AAH33065.  
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX PT useful for preventing, diagnosing and/or treating colorectal cancers.  
XX PS Claim 11; Page 6223-6224; 9803pp; English.  
XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where the  
XX CC proteins are collectively known as colon cancer antigens. The colon  
XX CC cancer antigens have cytostatic activity and can be used in gene therapy  
XX CC and vaccine production. N and P may be used in the prevention, diagnosis  
XX CC and treatment of diseases associated with inappropriate P expression. For  
XX CC example, N and P may be used to treat disorders associated with decreased  
XX CC expression by rectifying mutations or deletions in a patient's genome  
XX CC that affect the activity of P by expressing inactive proteins or to  
XX CC supplement the patient's own production of P. Additionally, N may be used  
XX CC to produce the colon cancer-associated P, by inserting the nucleic acids  
XX CC into a host cell and culturing the cell to express the proteins. N and P  
XX CC can be used in the prevention, diagnosis and treatment of colorectal  
XX CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent  
XX CC sequences used in the exemplification of the present invention. N.B.  
XX CC Pages 666 to 682 and page 7053 of the sequence listing were missing at  
XX CC time of publication, meaning no sequences are present for SEQ ID NO:1027  
XX CC to 1052, 7921 and 7922  
XX SQ Sequence 439 AA;  
Query Match 54.3%; Score 19; DB 4; Length 439;  
Best Local Similarity 30.8%; Pred. No. 4.7e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 4 YXXXXXXXRXSP 16  
DB 124 YKSTSSAFKSP 136  
RESULT 13  
ABB94042  
ID ABB94042 standard; protein; 461 AA.  
XX AC ABB94042;  
XX DT 06-JUN-2002 (first entry)

XX DE Human secreted protein SEQ ID NO: 85.  
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;  
XX KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX KW neurological disease; infection; human; secreted protein.  
XX OS Homo sapiens.  
XX PN WO200216389-A1.  
XX PD 28-FEB-2002.  
XX PF 17-JAN-2001; 2001WO-US001397.  
XX PR 18-AUG-2000; 2000US-0226281P.  
XX FA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR;  
XX PI Olsen HS, Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH;  
XX PI Fiscella M, Ni J;  
XX DR WPI; 2002-292054/33.  
XX DR N-PSDB; ABL92370.  
XX PT Isolated nucleic acids encoding 23 secreted proteins useful for the  
XX PT diagnosis and treatment of e.g. cancer, HIV infection, stroke and  
XX PT rheumatoid arthritis.  
XX PS Claim 11; Page 474-475; 517pp; English.  
XX CC Sequences ABB94017-94095 represent the amino acid sequences of 79 human  
XX CC secreted proteins encoded by the genes ABL92336-92393. The genes and  
XX CC proteins are useful for preventing, ameliorating or treating medical  
XX CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
XX CC a range of human tissues disclosed in the specification. The nucleic  
XX CC acids, proteins, antibodies and (ant)agonists are useful in the  
XX CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
XX CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)  
XX CC immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
XX CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
XX CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
XX CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
XX CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
XX CC infectious diseases such as viral, bacterial, fungal and parasitic  
XX CC infections  
XX SQ Sequence 461 AA;  
Query Match 54.3%; Score 19; DB 5; Length 461;  
Best Local Similarity 30.8%; Pred. No. 4.9e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 4 YXXXXXXXRXSP 16  
DB 146 YKSTSSAFKSP 158  
RESULT 14  
ABG65125  
ID ABG65125 standard; protein; 461 AA.  
XX AC ABG65125;  
XX DT 27-AUG-2002 (first entry)  
XX DE Human albumin fusion protein #1800.  
XX KW Albumin fusion protein; therapeutic protein X; human albumin; HA;

KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytostatic; antifertility; antiinflammatory; antiulcer;  
 KW immunomodulator; anti-HIV; antidiabetic; haenostatic; nootropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritic.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX WO200177137-A1.  
 XX 18-OCT-2001.  
 XX 12-APR-2001; 2001WO-US011988.  
 XX 12-APR-2000; 2000US-0229358P.  
 PR 25-APR-2000; 2000US-0199384P.  
 PR 21-DEC-2000; 2000US-0256931P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Haseltine WA;  
 PI WPI; 2002-010886/01.  
 DR New fusion protein for treating disease e.g. diabetes comprises an  
 PT albumin fused to a therapeutic protein.  
 XX Claim 1; Page 1770-1771; 2102pp; English.  
 XX The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA, also known as human serum  
 CC albumin, HSA). The proteins are useful for treating a disease or disorder  
 CC that may be modulated by therapeutic protein X. The albumin extends the  
 CC shelf-life of protein X, and may increase its biological in vitro/in vivo  
 CC activity. The protein is useful for treating and diagnosing disorders  
 CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
 CC disease, ulcerative colitis), immune disorders (e.g. acquired  
 CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
 CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
 CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
 CC schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
 CC ABG63326-ABG65518 represent albumin fusion proteins of the invention  
 XX Sequence 461 AA;  
 SQ  
 Query Match 54.3%; Score 19; DB 5; Length 461;  
 Best Local Similarity 30.8%; Pred. No. 4.9e+03;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 4 YXXXXXXXRXSP 16  
 Db 146 YYKSTSSAFRKSP 158  
 RESULT 15  
 AAB93892  
 ID AAB93892 standard; protein; 486 AA.  
 XX AAB93892;  
 XX 26-JUN-2001 (first entry)  
 XX Human protein sequence SEQ ID NO:13832.  
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KW Homo sapiens.  
 OS EP1074617-A2.  
 PN

PD 07-FEB-2001.  
 XX 28-JUL-2000; 2000EP-00116136.  
 XX 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX (HELI-) HELIX RES INST.  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX Claim 8; SEQ ID NO 13832; 2537pp + Sequence Listing; English.  
 PS The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX Sequence 486 AA;  
 SQ  
 Query Match 54.3%; Score 19; DB 4; Length 486;  
 Best Local Similarity 30.8%; Pred. No. 5.2e+03;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 4 YXXXXXXXRXSP 16  
 Db 171 YYKSTSSAFRKSP 183

Search completed: October 5, 2004, 16:06:36  
 Job time : 86.1918 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:00:17 ; Search time 21.6986 Seconds  
(without alignments)  
38.068 Million cell updates/sec

Title: US-09-973-473A-23  
Perfect score: 35  
Sequence: 1 XXXXXXXXXXRXSP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	51.4	217	3	US-08-341-018-48
2	18	51.4	217	3	US-08-341-018-50
3	18	51.4	217	3	US-08-470-335-205
4	18	51.4	217	3	US-08-470-335-205
5	18	51.4	217	4	US-08-467-602-399
6	18	51.4	217	4	US-08-467-602-413
7	18	51.4	238	4	US-09-252-991A-29254
8	18	51.4	239	3	US-08-470-335-219
9	18	51.4	239	3	US-08-470-335-219
10	18	51.4	273	1	US-08-221-750A-9
11	18	51.4	302	4	US-09-252-991A-27962
12	18	51.4	407	3	US-08-753-007A-6
13	18	51.4	407	3	US-09-398-496-6
14	18	51.4	416	1	US-08-252-995D-2
15	18	51.4	416	2	US-08-834-108-2
16	18	51.4	425	3	US-08-462-467B-16
17	18	51.4	427	4	US-08-467-602-189
18	18	51.4	430	4	US-08-467-602-187
19	18	51.4	436	4	US-08-467-602-201
20	18	51.4	439	4	US-08-467-602-198
21	18	51.4	450	4	US-08-467-602-195
22	18	51.4	459	4	US-08-467-602-204
23	18	51.4	464	1	US-08-252-995D-6
24	18	51.4	464	2	US-08-834-108-6
25	18	51.4	469	3	US-08-753-007A-8
26	18	51.4	469	3	US-09-398-496-8
27	18	51.4	474	4	US-08-467-602-190

28	18	51.4	477	4	US-08-467-602-185
29	18	51.4	479	1	US-07-923-724-2
30	18	51.4	479	2	US-08-609-426A-2
31	18	51.4	479	2	US-08-374-652C-4
32	18	51.4	483	4	US-08-467-602-202
33	18	51.4	486	4	US-08-467-602-199
34	18	51.4	497	4	US-08-467-602-193
35	18	51.4	506	4	US-08-467-602-205
36	18	51.4	560	3	US-08-341-018-58
37	18	51.4	560	3	US-08-470-335-194
38	18	51.4	560	4	US-08-470-339-194
39	18	51.4	560	4	US-08-467-602-388
40	18	51.4	569	4	US-08-467-602-237
41	18	51.4	572	4	US-08-467-602-235
42	18	51.4	578	4	US-08-467-602-249
43	18	51.4	581	4	US-08-467-602-246
44	18	51.4	592	4	US-08-467-602-243
45	18	51.4	601	4	US-08-467-602-252

ALIGNMENTS

RESULT 1  
US-08-341-018-48  
; Sequence 48, Application US/08341018A  
; Patent No. 6087323  
; GENERAL INFORMATION:  
; APPLICANT: Gwynne, David I.  
; APPLICANT: Mahanthappa, Nagesh K.  
; APPLICANT: Marchionni, Mark A.  
; APPLICANT: Bermingham-McDonogh, Olivia  
; APPLICANT: Goldin, Stanley M.  
; APPLICANT: McBurney, Robert N.  
; TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF  
; TITLE OF INVENTION: CELLULAR COMMUNICATION  
; FILE REFERENCE: 04585/041001  
; CURRENT APPLICATION NUMBER: US/08/341,018A  
; CURRENT FILING DATE: 1994-11-17  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Bos taurus  
US-08-341-018-48  
  
Query Match 51.4%; Score 18; DB 3; Length 217;  
Best Local Similarity 30.8%; Pred. No. 7.9e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
  
QY 4 YXXXXXXXRXSP 16  
DB 1 YVSAMTTTARMSP 13  
  
RESULT 2  
US-08-341-018-50  
; Sequence 50, Application US/08341018A  
; Patent No. 6087323  
; GENERAL INFORMATION:  
; APPLICANT: Gwynne, David I.  
; APPLICANT: Mahanthappa, Nagesh K.  
; APPLICANT: Marchionni, Mark A.  
; APPLICANT: Bermingham-McDonogh, Olivia  
; APPLICANT: Goldin, Stanley M.  
; APPLICANT: McBurney, Robert N.  
; TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF  
; TITLE OF INVENTION: CELLULAR COMMUNICATION  
; FILE REFERENCE: 04585/041001  
; CURRENT APPLICATION NUMBER: US/08/341,018A  
; CURRENT FILING DATE: 1994-11-17  
; NUMBER OF SEQ ID NOS: 87

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-341-018-50

Query Match          51.4%; Score 18; DB 3; Length 217;
Best Local Similarity 30.8%; Pred. No. 7.9e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
| | | | |
Db 1 YVSAMTTPARMSP 13

RESULT 3
US-08-470-335-205
; Sequence 205, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LOUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470,335F
; EARLIER FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/036,555
; EARLIER FILING DATE: 1993-03-24
; NUMBER OF SEQ ID NOS: 252
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 205
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Bos taurus
US-08-470-335-205

Query Match          51.4%; Score 18; DB 3; Length 217;
Best Local Similarity 30.8%; Pred. No. 7.9e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
| | | | |
Db 1 YVSAMTTPARMSP 13

RESULT 4
US-08-470-339-205
; Sequence 205, Application US/08470339C
; Patent No. 6232286
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LOUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470,339C
; EARLIER FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/036,555
; EARLIER FILING DATE: 1993-03-24
; EARLIER APPLICATION NUMBER: 07/940,389
; EARLIER FILING DATE: 1992-09-03

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 205
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Bos taurus
US-08-470-339-205

Query Match          51.4%; Score 18; DB 3; Length 217;
Best Local Similarity 30.8%; Pred. No. 7.9e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
| | | | |
Db 1 YVSAMTTPARMSP 13

RESULT 5
US-08-467-602-399
; Sequence 399, Application US/08467602C
; Patent No. 6444642
; GENERAL INFORMATION:
; APPLICANT: SKLAR, ROBERT
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: GWYNN, DAVID I.
; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 04585/028003
; CURRENT APPLICATION NUMBER: US/08/467,602C
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/209,204
; EARLIER FILING DATE: 1994-03-08
; EARLIER APPLICATION NUMBER: 08/059,022
; EARLIER FILING DATE: 1993-05-06
; NUMBER OF SEQ ID NOS: 420
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 399
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Bos taurus
US-08-467-602-399

Query Match          51.4%; Score 18; DB 4; Length 217;
Best Local Similarity 30.8%; Pred. No. 7.9e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
| | | | |
Db 1 YVSAMTTPARMSP 13

RESULT 6
US-08-467-602-413
; Sequence 413, Application US/08467602C
; Patent No. 6444642
; GENERAL INFORMATION:
; APPLICANT: SKLAR, ROBERT
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: GWYNN, DAVID I.
; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: 04585/028003
; CURRENT APPLICATION NUMBER: US/08/467,602C
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/209,204
; EARLIER FILING DATE: 1994-03-08
; EARLIER APPLICATION NUMBER: 08/059,022
```

```

; EARLIER FILING DATE: 1993-05-06
; NUMBER OF SEQ ID NOS: 420
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 413
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-467-602-413

Query Match      51.4%; Score 18; DB 4; Length 217;
Best Local Similarity 30.8%; Pred. No. 7.9e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 1 YVSAMTTPARMSP 13

RESULT 7
US-09-252-991A-29254
; Sequence 29254, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1995-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29254
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29254

Query Match      51.4%; Score 18; DB 4; Length 238;
Best Local Similarity 30.8%; Pred. No. 8.7e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 197 YAGAGAWANRTSP 209

RESULT 8
US-08-470-335-219
; Sequence 219, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; FILE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470.335F
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/036.555
; EARLIER FILING DATE: 1993-03-24
; EARLIER APPLICATION NUMBER: 91 07566.3
; NUMBER OF SEQ ID NOS: 252
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 219
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-335-219

Query Match      51.4%; Score 18; DB 3; Length 239;
Best Local Similarity 30.8%; Pred. No. 8.7e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 1 YVSAMTTPARMSP 13

RESULT 9
US-08-470-339-219
; Sequence 219, Application US/08470339C
; Patent No. 6232286
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; FILE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470.339C
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/036.555
; EARLIER FILING DATE: 1993-03-24
; EARLIER APPLICATION NUMBER: 07/940.389
; EARLIER FILING DATE: 1992-09-03
; EARLIER APPLICATION NUMBER: 07/907.138
; EARLIER FILING DATE: 1992-06-30
; EARLIER APPLICATION NUMBER: 07/863.703
; EARLIER FILING DATE: 1992-04-03
; EARLIER APPLICATION NUMBER: 91 07566.3 GB
; EARLIER FILING DATE: 1999-04-10
; NUMBER OF SEQ ID NOS: 226
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 219
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-339-219

Query Match      51.4%; Score 18; DB 3; Length 239;
Best Local Similarity 30.8%; Pred. No. 8.7e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 1 YVSAMTTPARMSP 13

RESULT 10
US-08-221-750A-9
; Sequence 9, Application US/08221750A
; Patent No. 5843747
; GENERAL INFORMATION:
; APPLICANT: Baker, Steven M.
; APPLICANT: Deich, Robert A.
; TITLE OF INVENTION: Genes for the Export of Pertussis
; FILE OF INVENTION: Holotoxin
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173

```

```

; ORGANISM: Homo sapiens
US-08-470-335-219

Query Match      51.4%; Score 18; DB 3; Length 239;
Best Local Similarity 30.8%; Pred. No. 8.7e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 1 YVSAMTTPARMSP 13

RESULT 9
US-08-470-339-219
; Sequence 219, Application US/08470339C
; Patent No. 6232286
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROOBANT, PAUL
; APPLICANT: MINGHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; FILE OF INVENTION: PREPARATION AND USE
; FILE REFERENCE: 04585/00200B
; CURRENT APPLICATION NUMBER: US/08/470.339C
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/036.555
; EARLIER FILING DATE: 1993-03-24
; EARLIER APPLICATION NUMBER: 07/940.389
; EARLIER FILING DATE: 1992-09-03
; EARLIER APPLICATION NUMBER: 07/907.138
; EARLIER FILING DATE: 1992-06-30
; EARLIER APPLICATION NUMBER: 07/863.703
; EARLIER FILING DATE: 1992-04-03
; EARLIER APPLICATION NUMBER: 91 07566.3 GB
; EARLIER FILING DATE: 1999-04-10
; NUMBER OF SEQ ID NOS: 226
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 219
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-470-339-219

Query Match      51.4%; Score 18; DB 3; Length 239;
Best Local Similarity 30.8%; Pred. No. 8.7e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 1 YVSAMTTPARMSP 13

RESULT 10
US-08-221-750A-9
; Sequence 9, Application US/08221750A
; Patent No. 5843747
; GENERAL INFORMATION:
; APPLICANT: Baker, Steven M.
; APPLICANT: Deich, Robert A.
; TITLE OF INVENTION: Genes for the Export of Pertussis
; FILE OF INVENTION: Holotoxin
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/221,750A  
FILING DATE: 31-MAR-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/031,619  
FILING DATE: 15-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 33,542  
REFERENCE/DOCKET NUMBER: ACC93-01A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 273 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-221-750A-9

Query Match 51.4%; Score 18; DB 1; Length 273;  
Best Local Similarity 30.8%; Pred. No. 9.9e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
Db 149 YPAAPQASRASP 161

## RESULT 11

US-09-252-991A-27962  
Sequence 27962, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 27962  
LENGTH: 302  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27962

Query Match 51.4%; Score 18; DB 4; Length 302;  
Best Local Similarity 30.8%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
Db 269 YTEAATIRIRISP 281

## RESULT 12

US-08-753-007A-6  
Sequence 6, Application US/08753007A  
Patent No. 6074841  
GENERAL INFORMATION:  
APPLICANT: Gearing, David P.

APPLICANT: Busfield, Samantha J.  
TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/753,007A  
FILING DATE: 19-NOV-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/699,591  
FILING DATE: 19-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, J. Peter  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07334/022001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 407 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-753-007A-6

Query Match 51.4%; Score 18; DB 3; Length 407;  
Best Local Similarity 30.8%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
Db 359 YVSALTTPARLSP 371

## RESULT 13

US-09-398-496-6  
Sequence 6, Application US/09398496  
Patent No. 6133423  
GENERAL INFORMATION:  
APPLICANT: Gearing, David P.  
APPLICANT: Busfield, Samantha J.  
TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/398,496  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/753,007  
FILING DATE: 19-NOV-1996  
APPLICATION NUMBER: 08/699,591  
FILING DATE: 19-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, J. Peter  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07334/022001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 407 amino acids  
TYPE: amino acid  
STRAINEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-09-398-496-6

Query Match 51.4%; Score 18; DB 3; Length 407;  
Best Local Similarity 30.8%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
DB 359 YVSALTTPARLSP 371  
RESULT 14  
US-09-252-995D-2  
Sequence 2, Application US/08252995D  
Patent No. 5650501  
GENERAL INFORMATION:  
APPLICANT: Dennis, James W  
APPLICANT: Heffernan, Mike  
APPLICANT: Fode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/252,995D  
FILING DATE: 02-JUN-1994  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurdydyk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

Query Match 51.4%; Score 18; DB 3; Length 407;  
Best Local Similarity 30.8%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

MOLECULE TYPE: protein  
US-08-252-995D-2  
Query Match 51.4%; Score 18; DB 1; Length 416;  
Best Local Similarity 30.8%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
DB 372 YLRRAHSSDRASP 384

RESULT 15  
US-08-834-108-2  
Sequence 2, Application US/08834108  
Patent No. 5976893  
GENERAL INFORMATION:  
APPLICANT: Dennis, James W  
APPLICANT: Heffernan, Mike  
APPLICANT: Fode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/834,108  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurdydyk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-834-108-2

Query Match 51.4%; Score 18; DB 2; Length 416;  
Best Local Similarity 30.8%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
DB 372 YLRRAHSSDRASP 384

Search completed: October 5, 2004, 16:15:39  
Job time : 22.6986 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:12:48 ; Search time 77.1507 Seconds  
(without alignments)  
66.737 Million cell updates/sec

Title: US-09-973-473A-23

Perfect score: 35

Sequence: 1 XXXYXXXXXXRXSP 16

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	60.0	617	15	US-10-369-493-3967
2	20	57.1	423	12	US-10-424-599-172660
3	20	57.1	431	12	US-10-425-114-36716
4	20	57.1	639	16	US-10-437-963-135098
5	20	57.1	1653	16	US-10-437-963-144523
6	20	57.1	1981	16	US-10-437-963-144521
7	19	54.3	146	16	US-10-767-701-38973
8	19	54.3	191	11	US-09-833-245-1875
9	19	54.3	193	11	US-09-833-245-1873
10	19	54.3	248	16	US-10-437-963-119735
11	19	54.3	382	9	US-09-801-368-216
12	19	54.3	439	14	US-10-106-698-4408
13	19	54.3	461	11	US-09-833-245-1874
14	19	54.3	486	11	US-10-308-448-25
15	19	54.3	486	16	US-10-343-593-16

16	19	54.3	728	12	US-10-282-122A-50474	Sequence 50474, A
17	19	54.3	1090	16	US-10-437-963-119369	Sequence 119369, A
18	19	54.3	1323	14	US-10-195-144-81	Sequence 81, Appl
19	19	54.3	1323	15	US-10-345-072-81	Sequence 81, Appl
20	19	54.3	2025	16	US-10-343-710-116	Sequence 116, Appl
21	19	54.3	6231	16	US-10-343-710-73	Sequence 73, Appl
22	19	54.3	6238	16	US-10-343-710-71	Sequence 71, Appl
23	18	51.4	73	16	US-10-437-963-126092	Sequence 126092, A
24	18	51.4	105	16	US-10-767-701-57483	Sequence 57483, A
25	18	51.4	163	16	US-10-767-701-34974	Sequence 34974, A
26	18	51.4	190	12	US-10-425-114-49619	Sequence 49619, A
27	18	51.4	198	12	US-10-425-114-59193	Sequence 59193, A
28	18	51.4	210	14	US-10-029-386-33437	Sequence 33437, A
29	18	51.4	224	12	US-10-425-114-50891	Sequence 50891, A
30	18	51.4	275	12	US-10-425-114-43743	Sequence 43743, A
31	18	51.4	344	9	US-09-795-668-38	Sequence 38, Appl
32	18	51.4	344	9	US-09-795-668-38	Sequence 38, Appl
33	18	51.4	344	9	US-09-946-807-38	Sequence 38, Appl
34	18	51.4	353	12	US-10-282-122A-61551	Sequence 61551, A
35	18	51.4	374	16	US-10-313-972-19	Sequence 19, Appl
36	18	51.4	375	16	US-10-313-972-23	Sequence 23, Appl
37	18	51.4	407	13	US-10-096-241-6	Sequence 6, Appl
38	18	51.4	425	10	US-09-742-153-16	Sequence 16, Appl
39	18	51.4	427	12	US-10-424-599-172666	Sequence 172666, A
40	18	51.4	447	16	US-10-437-963-117396	Sequence 117396, A
41	18	51.4	469	13	US-10-096-241-8	Sequence 8, Appl
42	18	51.4	530	12	US-10-424-599-247817	Sequence 247817, A
43	18	51.4	548	9	US-09-795-668-29	Sequence 29, Appl
44	18	51.4	548	9	US-09-795-668-29	Sequence 29, Appl
45	18	51.4	548	9	US-09-946-807-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1

US-10-369-493-3967  
; Sequence 3967, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3967  
; LENGTH: 617  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
; NAME/KEY: unsure  
; LOCATION: (1)..(617)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-3967

Query Match 60.0%; Score 21; DB 15; Length 617;  
Best Local Similarity 30.8%; Pred. No. 1.7e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXRXSP 16

Db 4 YTASTGSSLRASP 16

RESULT 2

US-10-424-599-172660

; Sequence 172660, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 172660

; LENGTH: 423

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_126928C.1.pep

US-10-424-599-172660

Query Match 57.1%; Score 20; DB 12; Length 423;

Best Local Similarity 30.8%; Pred. No. 2.4e+03;

Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXXSP 16

Db 368 YKGSVTATSRTP 380

RESULT 3

US-10-425-114-36716

; Sequence 36716, Application US/10425114

; Publication No. US2004003488A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 36716

; LENGTH: 431

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: 700756578\_FLI.pep

US-10-425-114-36716

Query Match 57.1%; Score 20; DB 12; Length 431;

Best Local Similarity 30.8%; Pred. No. 2.4e+03;

Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXXSP 16

Db 376 YKGSVTATSRTP 388

RESULT 4

US-10-437-963-135098

; Sequence 135098, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 135098

; LENGTH: 639

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_36806C.1.pep

US-10-437-963-135098

Query Match 57.1%; Score 20; DB 16; Length 639;

Best Local Similarity 30.8%; Pred. No. 3.4e+03;

Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXXSP 16

Db 79 YPDAAAAAARGSP 91

RESULT 5

US-10-437-963-144523

; Sequence 144523, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 144523

; LENGTH: 1653

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(1653)

; OTHER INFORMATION: unsure at all Xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_45328C.1.pep

US-10-437-963-144523

Query Match 57.1%; Score 20; DB 16; Length 1653;

Best Local Similarity 30.8%; Pred. No. 7.7e+03;

Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXXSP 16

Db 191 YATEGASSSRVSP 203

RESULT 6

US-10-437-963-144521

; Sequence 144521, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:



; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 144521  
; LENGTH: 1981  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_45326C.1.pep  
US-10-437-963-144521

Query Match 57.1%; Score 20; DB 16; Length 1981;  
Best Local Similarity 30.8%; Pred. No. 9e+03; 9; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16  
| | | | |  
Db 191 YATEGASSRVSP 203

RESULT 7  
US-10-767-701-38973  
; Sequence 38973, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53235)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 38973  
; LENGTH: 146  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C82161\_1.pep  
US-10-767-701-38973

Query Match 54.3%; Score 19; DB 16; Length 146;  
Best Local Similarity 30.8%; Pred. No. 1.8e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16  
| | | | |  
Db 38 YGSPRSASRVSP 50

RESULT 8  
US-09-833-245-1875  
; Sequence 1875, Application US/09833245  
; Publication No. US20040010134A1  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: US/09/833,245  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/229, 358

; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1875  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-833-245-1875

Query Match 54.3%; Score 19; DB 11; Length 191;  
Best Local Similarity 30.8%; Pred. No. 2.3e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16  
| | | | |  
Db 171 YYKSTSTFRKSP 183

RESULT 9  
US-09-833-245-1873  
; Sequence 1873, Application US/09833245  
; Publication No. US20040010134A1  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: US/09/833,245  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/229, 358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1873  
; LENGTH: 193  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (53)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-833-245-1873

Query Match 54.3%; Score 19; DB 11; Length 193;  
Best Local Similarity 30.8%; Pred. No. 2.3e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16  
| | | | |  
Db 171 YYKSTTSFAFRKSP 183

RESULT 10  
US-10-437-963-119735  
; Sequence 119735, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; PRIOR FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 119735
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_22923C.1.pap
US-10-437-963-119735

Query Match          54.3%; Score 19; DB 16; Length 248;
Best Local Similarity 30.8%; Pred. No. 2.8e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
Db 180 YYAHSYSHRASP 192

RESULT 11
US-09-801-368-216
; Sequence 216, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Bueby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: NO. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR FILING DATE: 2000-01-19
; PRIOR FILING DATE: 2000-01-19
; PRIOR FILING DATE: 2000-01-19
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 216
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-216

Query Match          54.3%; Score 19; DB 9; Length 382;
Best Local Similarity 30.8%; Pred. No. 4.1e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
Db 220 YISNSSQSRQSP 232

RESULT 12
US-10-106-698-4408
; Sequence 4408, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA00591
```

```
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4408
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-4408

Query Match          54.3%; Score 19; DB 14; Length 439;
Best Local Similarity 30.8%; Pred. No. 4.7e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
Db 124 YKSTSSAFRKSP 136

RESULT 13
US-09-833-245-1874
; Sequence 1874, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PFS46PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1874
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (168)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (169)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (171)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (178)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (442)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-1874

Query Match          54.3%; Score 19; DB 11; Length 461;
Best Local Similarity 30.8%; Pred. No. 4.9e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16
```

```
Db      146 YYKSTSSAFRKSP 158
RESULT 14
US-10-308-448-25
; Sequence 25, Application US/10308448
; Publication No. US20030170743A1
; GENERAL INFORMATION:
; APPLICANT: Bristo-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE DEVELOPMENT
; FILE REFERENCE: D0189
; CURRENT APPLICATION NUMBER: US/10/308,448
; CURRENT FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: 60/337,417
; PRIOR FILING DATE: 2001-12-03
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-308-448-25

Query Match      54.3%; Score 19; DB 14; Length 486;
Best Local Similarity 30.8%; Pred. No. 5.1e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy      4 YXXXXXXXRXSP 16
Db      171 YYKSTSSAFRKSP 183

RESULT 15
US-10-343-593-16
; Sequence 16, Application US/10343593
; Publication No. US20040110259A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; BAUGHN, Mariah R.;
; APPLICANT: BRUNS, Christopher M.; DAS, Debopriya;
; APPLICANT: DELEGANE, Augusto M.; DING, Li;
; APPLICANT: ELLIOT, Vicki S.; GANDHI, Aameena R.;
; APPLICANT: GRIFFIN, Jennifer A.; HAPALIA, April J.A.;
; APPLICANT: KHAN, Farrah A.; LAL, Preeti G.;
; APPLICANT: LEE, Sally; LU, Dyung Aina M.;
; APPLICANT: LU, Yan; ARVIZU, Chandra S.;
; APPLICANT: RAMKOMAR, Jayalaxmi; RING, Huijun Z.;
; APPLICANT: SANJANWALA, Madhusudan M.; TANG, Y. Tom;
; APPLICANT: THANGAVELU, Kavitha; THORNTON, Michael;
; APPLICANT: TRIBOULEY, Catherine M.; CHAWLA, Narinder K.;
; APPLICANT: WARREN, Bridget A.; YANG, Junming;
; APPLICANT: YAO, Monique G.; YUE, Henry
; TITLE OF INVENTION: DRUG METABOLIZING ENZYMES
; FILE REFERENCE: PI-0185 USN
; CURRENT APPLICATION NUMBER: US/10/343,593
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US 60/223,055
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,728
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/226,440
; 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/228,067
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/230,063
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 60/232,244
; PRIOR FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: US 60/234,269
; PRIOR FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PERL Program
```

```
; SEQ ID NO 16
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 1618256CD1
US-10-343-593-16

Query Match      54.3%; Score 19; DB 16; Length 486;
Best Local Similarity 30.8%; Pred. No. 5.1e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy      4 YXXXXXXXRXSP 16
Db      171 YYKSTSSAFRKSP 183

Search completed: October 5, 2004, 16:47:05
Job time : 79.1507 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:58:01 ; Search time 19.5068 Seconds  
(without alignments)  
78.899 Million cell updates/sec

Title: US-09-973-473A-23

Perfect score: 35  
Sequence: 1 XXXYXXXXXXRXSP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	60.0	589	2 T49653	related to SLV41 p
2	20	57.1	115	2 F72570	hypothetical prote
3	19	54.3	104	2 S56778	probable membrane
4	19	54.3	382	2 S51881	MSNI protein - yea
5	19	54.3	455	2 T50426	hypothetical prote
6	19	54.3	468	2 B70932	probable PPS prote
7	19	54.3	567	2 A84748	hypothetical prote
8	19	54.3	2207	1 GNNY1P	genome polyprotein
9	19	54.3	2209	1 GNNY2P	genome polyprotein
10	19	54.3	2209	1 GNNY3P	genome polyprotein
11	18	51.4	81	2 AC2097	hypothetical prote
12	18	51.4	150	2 S03872	hypothetical prote
13	18	51.4	265	2 B83395	probable enoyl-CoA
14	18	51.4	266	2 T36439	hypothetical prote
15	18	51.4	273	2 D47301	VirB9 homolog - Bo
16	18	51.4	346	2 T70874	probable PPS prote
17	18	51.4	350	2 I38403	neu differentiation
18	18	51.4	354	2 A12889	lysyl-tRNA synthet
19	18	51.4	354	2 F97665	lysyl-tRNA synthet
20	18	51.4	363	2 T17543	probable D-lactate
21	18	51.4	365	2 T42954	hypothetical prote
22	18	51.4	394	2 G70881	probable PPS prote
23	18	51.4	396	1 AJMZRB	argininosuccinate
24	18	51.4	402	2 T40882	probable PPS prote
25	18	51.4	425	2 T41683	hypothetical prote
26	18	51.4	456	2 T40416	hypothetical prote
27	18	51.4	465	2 B55748	protein kinase (EC
28	18	51.4	479	1 JN0715	3-phytase (EC 3.1.
29	18	51.4	479	1 JN0890	acid phosphatase (

30	18	51.4	490	1 G59282	hypothetical prote
31	18	51.4	602	2 A45769	acetylcholine rece
32	18	51.4	636	2 I61718	neu differentiation
33	18	51.4	637	2 C43273	heregulin precurs
34	18	51.4	639	2 I61719	neu differentiation
35	18	51.4	640	2 A43273	heregulin precurs
36	18	51.4	645	2 B43273	heregulin, splice
37	18	51.4	662	2 I61722	neu differentiation
38	18	51.4	850	2 JCS702	Erbb kinase activa
39	18	51.4	860	2 JCS702	Erbb kinase activa
40	18	51.4	868	2 JCS701	Erbb kinase activa
41	18	51.4	925	2 A55748	protein kinase (EC
42	17	48.6	101	2 H90223	ribosomal protein
43	17	48.6	155	2 T22815	hypothetical prote
44	17	48.6	162	2 G4351	hypothetical prote
45	17	48.6	171	2 B87018	probable membrane

ALIGNMENTS

RESULT 1

T49653  
related to SLV41 protein [imported] - Neurospora crassa  
N:Alternate names: protein B8B20.70  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
R:Schulte, U.; Aign, V.; Hobeisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49653  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-589 <SCH>  
A:Cross-references: ENBL:AL355933; GSPDB:GN00116; NCSP:B8B20.70  
A:Experimental source: BAC clone B8B20; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B8B20.70  
A:Map position: 6  
A:Introns: 126/3

Query Match 60.0%; Score 21; DB 2; Length 589;  
Best Local Similarity 30.8%; Pred. NO. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
DB 4 YTAGTSSLRASP 16

RESULT 2

F72570  
hypothetical protein APE1847 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: F72570  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-115 <KAW>  
A:Cross-references: DBJ:AP000062; NID:G5105244; PIDN:BAA80851.1; PID:d1044637; PID:G510:  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1847  
C:Superfamily: Aeropyrum pernix hypothetical protein APE1847

Query Match 57.1%; Score 20; DB 2; Length 115;

Best Local Similarity 30.8%; Pred. No. 50;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
| | | | |  
Db 80 YASTSTFTTRPSP 92

RESULT 3  
S56778  
probable membrane protein YJL007c - yeast (*Saccharomyces cerevisiae*)  
N;Alternate names: hypothetical protein J1379  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 19-Apr-2002  
A;Accession: S56778  
R;To Van, D.; Perea, J.; Jacq, C.  
submitted to the Protein Sequence Database, September 1995  
A;Reference number: S56776  
A;Accession: S56778  
A;Molecule type: DNA  
A;Residues: 1-104 <DEH>  
A;Cross-references: EMBL:Z49282; NID:g1006717; PID:g1006718; MIPS:YJL007c  
C;Genetics:  
A;Gene: MIPS:YJL007c  
A;Cross-references: SGD:S0003544  
A;Map position: 10L  
C;Superfamily: *Saccharomyces cerevisiae* probable membrane protein YJL007c  
C;Keywords: transmembrane protein

Query Match 54.3%; Score 19; DB 2; Length 104;  
Best Local Similarity 30.8%; Pred. No. 97;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
| | | | |  
Db 50 YSTGYTGHTRSSP 62

RESULT 4  
S51881  
MSN1 protein - yeast (*Saccharomyces cerevisiae*)  
N;Alternate names: FUP1 protein; protein HRB382; protein O0713; protein YOL116w  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 05-May-1995 #sequence\_revision 03-Aug-1995 #text\_change 21-Jul-2000  
A;Accession: S51881; S12325; S59157; S22771; S56812  
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, P.  
submitted to the EMBL Data Library, January 1995  
A;Description: Sequence analysis of a 44kb DNA fragment of yeast chromosome XV including and a Delta.

A;Reference number: S51848  
A;Accession: S51881  
A;Molecule type: DNA  
A;Residues: 1-382 <VAN>  
A;Cross-references: EMBL:Z48149; NID:g663234; PID:g663236  
R;Estreich, F.; Carlson, M.  
Nucleic Acids Res 18, 6959-6964, 1990  
A;Title: Increased dosage of the MSN1 gene restores invertase expression in yeast mutant  
A;Reference number: S12325; MUID:9108271; PMID:2263457  
A;Accession: S12325

A;Molecule type: DNA  
A;Residues: 1-359, 'Y', 361-382 <EST>  
A;Cross-references: EMBL:X54324; NID:g3999; PIDN:CAA38222.1; PID:g4000  
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, P.  
Yeast 11, 1069-1075, 1995  
A;Title: Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including the a delta element.

A;Reference number: S59156; MUID:96076631; PMID:7502582  
A;Accession: S59157  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-382 <VAN>  
A;Cross-references: EMBL:Z48149; NID:g663234; PIDN:CAA88144.1; PID:g663236  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995

R;Eide, D.; Guarente, L.  
J. Gen. Microbiol 138, 347-354, 1992  
A;Title: Increased dosage of a transcriptional activator gene enhances iron-limited growth  
A;Reference number: S22771; MUID:9222698; PMID:1564445  
A;Accession: S22771  
A;Molecule type: DNA  
A;Residues: 1-40 <EID>  
R;Durand, P.; Hilger, P.; Portetelle, D.; Vandenbol, M.  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S66791  
A;Accession: S66812  
A;Molecule type: DNA  
A;Residues: 1-382 <DUR>  
A;Cross-references: EMBL:Z74858; NID:g1419988; PID:g252298; PID:g1419989; MIPS:YOL116w  
A;Experimental source: strain S288C

C;Genetics:  
A;Gene: SGD:MSN1; FUP1  
A;Cross-references: SGD:S0005476; MIPS:YOL116w  
A;Map position: 15L  
C;Keywords: DNA binding; nucleus

Query Match 54.3%; Score 19; DB 2; Length 382;  
Best Local Similarity 30.8%; Pred. No. 3.4e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
| | | | |  
Db 220 YISSNSSQSRQSP 232

RESULT 5  
T50426  
hypothetical protein SPCC24B10.19c with probable coiled-coil region [imported] - fission  
C;Species: *Schizosaccharomyces pombe*  
C;Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
A;Accession: T50426  
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, February 2000  
A;Reference number: Z25039

A;Accession: T50426  
A;Status: preliminary; translated from GB/EMBL/DBBJ  
A;Molecule type: DNA  
A;Residues: 1-455 <SEE>  
A;Cross-references: EMBL:AL157991; PIDN:CAB76228.1; GSPDB:GN00068; SPDB:SPCC24B10.19c  
A;Experimental source: strain 972h(-); cosmid c24B10  
C;Genetics:  
A;Gene: SPDB:SPCC24B10.19c  
A;Map position: 3  
A;Introns: 16/1

Query Match 54.3%; Score 19; DB 2; Length 455;  
Best Local Similarity 23.1%; Pred. No. 4e+02;  
Matches 3; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
| | | | |  
Db 280 YLTSTTTSTKSP 292

RESULT 6  
B70932  
probable PPE protein - *Mycobacterium tuberculosis* (strain H37RV)  
C;Species: *Mycobacterium tuberculosis*  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
A;Accession: B70932  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: B70932

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-468 <COL>  
A:Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN:CAA17730.1; PID:e125462  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: PPE

Query Match 54.3%; Score 19; DB 2; Length 468;  
Best Local Similarity 23.1%; Pred. No. 4.1e+02;  
Matches 3; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16  
Db 154 YASSATASRLTP 166

RESULT 7  
A84748  
Hypothetical protein At2g33640 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: A84748  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: A84748  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-567 <STO>  
A:Cross-references: GB:AB002093; NID:G2459444; PIDN:AB80679.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g33640  
A:Map position: 2

Query Match 54.3%; Score 19; DB 2; Length 567;  
Best Local Similarity 30.8%; Pred. No. 4.9e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16  
Db 412 YKTSNVSGRSP 424

RESULT 8  
GNVY1P  
Genome polyprotein (version 1) - human poliovirus 1 (strain Mahoney)  
N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core  
tein P3-1b; protein P3-2; RNA-directed RNA polymerase (EC 2.7.7.48) P3-4b  
C:Species: human poliovirus 1  
C:Date: 17-Dec-1982 #sequence\_revision 17-Dec-1982 #text\_change 16-Jul-1999  
R:Kitamura, N.; Semler, B.L.; Rothberg, P.G.; Larsen, G.R.; Adler, C.J.; Dorner, A.J.; E  
Nature 291, 547-553, 1981  
A:Title: Primary structure, gene organization and polypeptide expression of poliovirus R  
A:Reference number: A93259; MUID:81220953; PMID:6264310  
A:Accession: A93258  
A:Molecule type: genomic RNA  
A:Residues: 1-2207 <KIT1>  
A:Cross-references: GB:V01148; NID:G61236; PIDN:CAA24446.1; PID:G61237  
A:Note: the amino acid sequence of VPg (residues 1543-1564) was also determined and agre  
R:Kitamura, N.; Adler, C.J.; Rothberg, P.G.; Martinko, J.; Nathenson, S.G.; Wimmer, E.  
Cell 21, 295-302, 1980  
A:Title: The genome-linked protein of picornaviruses. VII. Genetic mapping of poliovirus  
A:Reference number: A90800; MUID:81001866; PMID:6250717  
A:Accession: A90800  
A:Molecule type: genomic RNA  
A:Residues: 1539-1574 <KIT2>  
A:Note: the amino end of VPg corresponds to residue 1543; a choice between the two poten  
in obtained by radiochemical microsequence analysis agrees with that predicted by the vi

R:Rothberg, P.G.; Harris, T.J.; Nomoto, A.; Wimmer, E.  
Proc. Natl. Acad. Sci. U.S.A. 75, 4868-4872, 1978  
A:Title: O4-(5'-Uridyl)tyrosine is the bond between the genome-linked protein and the f  
A:Reference number: A30637; MUID:79116223; PMID:217003  
A:Contents: annotation; chemical characterization  
C:Comment: VPg is linked by Tyr-1545 to the uridylylate residue at the 5' end of the genome  
C:Comment: Coat proteins VP2 and VP3 and the RNA-directed RNA polymerase are related to t  
C:Superfamily: poliovirus genome polyprotein  
C:Keywords: genome-linked protein; nucleotidyltransferase; phosphoprotein; polyprotein  
F:2-69/Product: coat protein VP4 #status predicted <VP4>  
F:70-340/Product: coat protein VP2 #status predicted <VP2>  
F:341-578/Product: coat protein VP3 #status predicted <VP3>  
F:579-880/Product: coat protein VP1 #status predicted <VP1>  
F:881-1455/Product: core protein P2-3b #status predicted <P23>  
F:1030-1455/Product: core protein P2-5b #status predicted <P25>  
F:1127-1455/Product: core protein P2-X #status predicted <P2X>  
F:1456-2207/Product: protein P3-1b #status predicted <P31>  
F:1543-1564/Product: genome-linked protein VPg #status predicted <VPg>  
F:1565-2207/Product: protein P3-2 #status predicted <P32>  
F:1565-1746/Product: probable proteinase P3-7c #status predicted <P37>  
F:1747-2207/Product: RNA-directed RNA polymerase P3-4b #status predicted <P34>  
F:1545/Binding site: phosphoryl-RNA (Tyr) (covalent) #status experimental

Query Match 54.3%; Score 19; DB 1; Length 2207;  
Best Local Similarity 30.8%; Pred. No. 1.8e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16  
Db 1343 YVLA TSSSRISP 1355

RESULT 9  
GNVY2P  
Genome polyprotein (version 2) - human poliovirus 1 (strain Mahoney)  
C:Species: human poliovirus 1  
C:Date: 18-Dec-1981 #sequence\_revision 18-Dec-1981 #text\_change 21-Jul-2000  
R:Racaniello, V.R.; Baltimore, D.  
Proc. Natl. Acad. Sci. U.S.A. 78, 4887-4891, 1981  
A:Title: Molecular cloning of poliovirus cDNA and determination of the complete nucleotic  
A:Reference number: A03898; MUID:82060159; PMID:6272282  
A:Accession: A03898  
A:Molecule type: genomic RNA  
A:Residues: 1-2209 <RAC>  
A:Cross-references: GB:J02281; NID:G61252; PIDN:CAA24461.1; PID:G61253  
A:Note: the authors translated the codon TTC for residue 464 as Ser  
R:Ambros, V.; Baltimore, D.  
J. Biol. Chem. 253, 5263-5266, 1978  
A:Title: Protein is linked to the 5' end of poliovirus RNA by a phosphodiester linkage t  
A:Reference number: A30636; MUID:78218195; PMID:209034  
A:Contents: annotation; chemical characterization  
C:Superfamily: poliovirus genome polyprotein  
C:Keywords: genome-linked protein; nucleotidyltransferase; phosphoprotein; polyprotein  
F:2-69/Product: coat protein VP4 #status predicted <VP4>  
F:70-341/Product: coat protein VP2 #status predicted <VP2>  
F:342-579/Product: coat protein VP3 #status predicted <VP3>  
F:580-881/Product: coat protein VP1 #status predicted <VP1>  
F:882-1456/Product: core protein P2-3b #status predicted <P23>  
F:1031-1456/Product: core protein P2-5b #status predicted <P25>  
F:1128-1456/Product: core protein P2-X #status predicted <P2X>  
F:1457-2209/Product: protein P3-1b #status predicted <P31>  
F:1544-1565/Product: genome-linked protein VPg #status predicted <VPg>  
F:1566-2209/Product: protein P3-2 #status predicted <P32>  
F:1566-1748/Product: probable proteinase P3-7c #status predicted <P37>  
F:1749-2209/Product: RNA-directed RNA polymerase P3-4b #status predicted <P34>  
F:1546/Binding site: phosphoryl-RNA (Tyr) (covalent) #status experimental

Query Match 54.3%; Score 19; DB 1; Length 2209;  
Best Local Similarity 30.8%; Pred. No. 1.8e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16

Db 1344 YVLASTNSSRISP 1356

RESULT 10  
GNYP3P  
genome polyprotein - human poliovirus 1 (strain Sabin)  
C:Species: human poliovirus 1  
C:Date: 17-Dec-1982 #sequence\_revision 17-Dec-1982 #text\_change 16-Jul-1999  
C:Accession: A03899  
R:Nomoto, A.; Omata, T.; Toyoda, H.; Kuge, S.; Horie, H.; Kataoka, Y.; Genba, Y.; Nakano  
Proc. Natl. Acad. Sci. U.S.A. 79, 5793-5797, 1982  
A:Title: Complete nucleotide sequence of the attenuated poliovirus Sabin 1 strain genome  
A:Reference number: A03899; MUID:83299876; PMID:6310545  
A:Accession: A03899  
A:Molecule type: genomic RNA  
A:Residues: 1-2209 <NOW>  
A:Cross-references: GB:J02282; GB:J02285; GB:J02286; GB:V01133; NID:g61257; F  
A:Note: this virus is a live vaccine strain derived from the Mahoney strain by spontaneous  
C:Superfamily: poliovirus genome polyprotein  
C:Keywords: genome-linked protein; phosphoprotein; polyprotein  
F:2-69/Product: coat protein VP4 #status predicted <VP4>  
F:70-341/Product: coat protein VP2 #status predicted <VP2>  
F:342-579/Product: coat protein VP3 #status predicted <VP3>  
F:580-881/Product: coat protein VP1 #status predicted <VP1>  
F:882-1456/Product: core protein P2-3b #status predicted <P23>  
F:1031-1456/Product: core protein P2-5b #status predicted <P25>  
F:1128-1456/Product: core protein P2-X #status predicted <P2X>  
F:1457-2209/Product: protein P3-1b #status predicted <P31>  
F:1544-1565/Product: genome-linked protein VPg #status predicted <VPG>  
F:1566-2209/Product: protein P3-2 #status predicted <P32>  
F:1566-1748/Product: probable proteinase P3-7c #status predicted <P37>  
F:1749-2209/Product: RNA-directed RNA polymerase P3-4b #status predicted <P34>  
F:1546/Binding site: phosphoryl-RNA (Ty) (covalent) #status predicted

Query Match 54.3%; Score 19; DB 1; Length 2209;  
Best Local Similarity 30.8%; Pred. No. 1.8e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
Db 1344 YVLASTNSSRISP 1356

RESULT 11  
AC2097  
hypotheical protein asr2330 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AC2097  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2097  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-81 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA074029.1; PID:g17131422; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: asr2330

Query Match 51.4%; Score 18; DB 2; Length 81;  
Best Local Similarity 30.8%; Pred. No. 1.6e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
Db 61 YSQTGSKSRNSP 73

RESULT 12  
S09872  
hypotheical protein UL107 - human cytomegalovirus (strain AD169)  
C:Species: human cytomegalovirus, human herpesvirus 5  
A:Note: host Homo sapiens (man)  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 28-Jul-2000  
C:Accession: S09872  
R:Chae, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; F  
M.; Barrell, B.G.  
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
A:Reference number: S09749; MUID:90269039; PMID:2161319  
A:Accession: S09872  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-150 <CHE>  
A:Cross-references: EMBL:X17403; NID:g59591; PIDN:CAA5343.1; PID:e27308; PID:g1780886  
A:Note: this sequence was submitted to the EMBL Data Library, December 1989  
A:Note: this reading frame extends between two stop codons and does not begin with a star  
C:Superfamily: human cytomegalovirus hypothetical protein UL107

Query Match 51.4%; Score 18; DB 2; Length 150;  
Best Local Similarity 30.8%; Pred. No. 2.9e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
Db 4 YSITVYDHRISP 16

RESULT 13  
B83395  
probable enoyl-CoA hydratase/isomerase PA2013 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83395  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathos  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: B83395  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-265 <STO>  
A:Cross-references: GB:AE004627; GB:AE004091; NID:g9948007; PIDN:AAG05401.1; GSPDB:GN001;  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2013  
C:Superfamily: naphthoate synthase; enoyl-CoA hydratase homology

Query Match 51.4%; Score 18; DB 2; Length 265;  
Best Local Similarity 30.8%; Pred. No. 5e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
Db 232 YTEAIAIRIRISP 244

RESULT 14  
T36439  
hypotheical protein SCF43A.16 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T36439  
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1999  
A:Reference number: Z21598  
A:Accession: T36439  
A:Status: preliminary; translated from GB/EMBL/DBDJB



A:Molecule type: DNA  
A:Residues: 1-266 <SEE>  
A:Cross-references: EMBL:AL096837; PIDN:CAB48903.1; GSPDB:GN00070; SCOEDB:SCF43A.16  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCF43A.16

Query Match 51.4%; Score 18; DB 2; Length 266;  
Best Local Similarity 30.8%; Pred. No. 5.1e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXXSP 16  
| | | | |  
Db 56 YRAVMSDARRSP 68

RESULT 15  
D47301  
VirB9 homolog - Bordetella pertussis  
C:Species: Bordetella pertussis  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994  
C:Accession: D47301  
R:Weiss, A.A.; Johnson, F.D.; Burns, D.L.  
Proc. Natl. Acad. Sci. U.S.A. 90, 2970-2974, 1993  
A:Title: Molecular characterization of an operon required for pertussis toxin secretion.  
A:Reference number: A47301; MUID:93219406; PMID:8464913  
A:Contents: BP338  
A:Accession: D47301  
A>Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-273 <WEI>  
A:Note: sequence extracted from NCBI backbone (NCBIN:128775, NCBIP:128780)

Query Match 51.4%; Score 18; DB 2; Length 273;  
Best Local Similarity 30.8%; Pred. No. 5.2e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXXSP 16  
| | | | |  
Db 149 YPAAPQAASRAS 161

Search completed: October 5, 2004, 16:13:55  
Job time : 21.5068 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:36 ; Search time 11.1781 Seconds  
(without alignments)  
74.532 Million cell updates/sec

Title: US-09-973-473a-23

Perfect score: 35  
Sequence: 1 XXXYXXXXXXXRXSP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	19	54.3	104	1 YJA7 YEAST	P47080 saccharomyc
2	19	54.3	382	1 MSN1 YEAST	P22148 saccharomyc
3	19	54.3	2206	1 POLG POLIM	P03299 p genome po
4	19	54.3	2208	1 POLH POLIM	P03300 p genome po
5	19	54.3	2209	1 POLG POLIS	P03301 p genome po
6	18	51.4	150	1 ULA7 HCMVA	P16828 human cytom
7	18	51.4	353	1 OXAA MYCPA	Q917ml mycobacteri
8	18	51.4	394	1 ASSY METAC	Q8tny5 methanosarc
9	18	51.4	394	1 ASSY METWA	Q8gou5 methanosarc
10	18	51.4	396	1 ASSY METBA	P13257 methanosarc
11	18	51.4	479	1 PHYB ASPAW	P34755 aspergillus
12	18	51.4	479	1 PHYB ASPNG	P34754 aspergillus
13	18	51.4	501	1 GYG2 HUMAN	O15488 homo sapien
14	18	51.4	602	1 NRG1 CHICK	Q05199 gallus gall
15	18	51.4	639	1 NRG1 HUMAN	Q02297 h pro-neure
16	18	51.4	662	1 NRG1 RAT	P43322 r pro-neure
17	18	51.4	677	1 NRG1 XENLA	O93383 xenopus lae
18	18	51.4	756	1 NRG2 MOUSE	P56974 mus musculu
19	18	51.4	850	1 NRG2 HUMAN	O14511 homo sapien
20	18	51.4	868	1 NRG2 RAT	O35569 rattus norv
21	17	48.6	101	1 RU21 SULSO	Q9uxeo sulfolobus
22	17	48.6	162	1 Y415 METJA	Q57858 methanococc
23	17	48.6	261	1 YGCW ECOLI	P76633 escherichia
24	17	48.6	262	1 YCFZ ECOLI	P75961 escherichia
25	17	48.6	267	1 FPG THERH	O50606 thermus the
26	17	48.6	310	1 TP2E PYROC	Q977x4 pyrodictum
27	17	48.6	314	1 HEM3 BUCAL	P57651 buchnera ap
28	17	48.6	366	1 VGLM HSVSA	Q01017 herpesvirus
29	17	48.6	390	1 ASSY ARCFU	O28032 archaeoglob
30	17	48.6	396	1 ASSY STRMU	Q8cwz0 streptococc
31	17	48.6	398	1 ASSY STRR6	Q8dri5 streptococc
32	17	48.6	401	1 ASSY CHLTE	Q8kde0 chlorobium
33	17	48.6	477	1 MYPH HUMAN	Q13203 homo sapien

RESULT 1

YJA7 YEAST  
ID YJA7 YEAST STANDARD; PRT; 104 AA.

AC P47080;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Hypothetical 11.9 kDa protein in CTR8-CTR2 intergenic region.  
GN YJL007C OR J1379.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA To Van D., Perea J., Jacq C.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z49282; CAA89298.1; -.  
DR PIR; S56778; S56778.  
DR GerOnline; 141623; -.  
DR SGD; S0003544; YJL007C.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 72 92 POTENTIAL.  
SQ SEQUENCE 104 AA; 11856 MW; DEBA235CCAD26770 CRC64;

Query Match 54.3%; Score 19; DB 1; Length 104;  
Best Local Similarity 30.8%; Pred. No. 42;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
DB 50 YSTGVTGHTRSSP 62

RESULT 2

MSN1 YEAST  
ID MSN1 YEAST STANDARD; PRT; 382 AA.

AC P22148; Q12227; Q9URF6;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE MSN1 protein (Multicopy suppressor of SNF1 protein 1).  
GN MSN1 OR FUP1 OR PHD2 OR YOL116W OR HRB382.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=91108271; PubMed=2263457;
RA Estruch F., Carlson M.;
RT "Increased dosage of the MSN1 gene restores invertase expression in
RL yeast mutants defective in the SNF1 protein kinase.";
RL Nucleic Acids Res. 18:6959-6964(1990).
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=92226698; PubMed=1564445;
RA Eide D., Guarente L.;
RT "Increased dosage of a transcriptional activator gene enhances iron-
RL limited growth of Saccharomyces cerevisiae.";
RL J. Gen. Microbiol. 138:347-354(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96076631; PubMed=7502582;
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT "Sequence analysis of a 44 Kb DNA fragment of yeast chromosome XV
RL including the Tyl-H3 retrotransposon, the sufi(+)- frameshift
RT suppressor gene for tRNA-Gly, the Yeast transfer RNA-Thr-Ia and a
RL delta element.";
RL Yeast 11:1069-1075(1995).
CC -I- FUNCTION: May function as a transcriptional activator. Increased
CC dosage of MSN1 restores invertase expression in yeast mutants
CC defective in the SNF1 protein kinase, and men1 disruption reduced
CC derepression of invertase in the wild type. May affect SUC2
CC expression. Expression of MSN1 enhances growth in iron-limiting
CC conditions.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54324; CAA38222.1; -.
CC ENBL; Z48149; CAA88144.1; -.
CC ENBL; Z74858; CAA99135.1; -.
CC PIR; S51881; S51881.
CC Germonline; 143538; -.
CC TRANSFAC; T03501; -.
CC SGD; S0005476; MSN1.
CC DR GO; GO:0005634; C:nucleus; IDA.
CC DR GO; GO:0006972; P:hyperosmotic response; IGI.
CC KW Nuclear protein; DNA-binding; Activator; Transcription regulation.
CC FT DOMAIN 12 26 LEUCINE-ZIPPER.
CC FT DOMAIN 266 271 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT CONFLICT 360 360 D -> Y (IN REF. 1).
CC SQ SEQUENCE 382 AA; 43060 MW; 2E09FD17A3A5B715 CRC64;
CC -----
CC Query Match 54.3%; Score 19; DB 1; Length 382;
CC Best Local Similarity 30.8%; Pred. No. 1.7e+02;
CC Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
CC -----
CC Qy 4 YXXXXXXXRXSP 16
CC 220 YISSNSQSRQSP 232
CC -----
CC RESULT 3
CC ID POLG POLIM STANDARD; PRT; 2206 AA.
CC AC P03299;
CC DT 21-JUL-1986 (Rel. 01, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Genome polypeptide [Contains: Coat protein VP4 (PIA), Coat protein VP2
CC (PIB); Coat protein VP3 (PIC); Coat protein VP1 (PID); Picornain 2A
CC -----
DE DE
DE DE
DE DE
OS OS
OC OC
OC OC
OC NCBI_TaxID=12081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81220953; PubMed=6264310;
RA Kitamura N., Semler B.L., Rothberg P.G., Larsen G.R., Adler C.J.,
RA Dorner A.J., Emili E.A., Hanecak R., Lee J.J., van der Werf S.,
RA Anderson C.W., Wimmer E.;
RT "Primary structure, gene organization and polypeptide expression of
RT poliovirus RNA.";
RL Nature 291:547-553(1981).
RN [2]
RP SEQUENCE OF 1538-1573 FROM N.A.
RX MEDLINE=81001866; PubMed=6250717;
RA Kitamura N., Adler C.J., Rothberg P.G., Martinko J., Nathenson S.G.,
RA Wimmer E.;
RT "The genome-linked protein of picornaviruses. VII. Genetic mapping of
RT poliovirus VPg by protein and RNA sequence studies.";
RL Cell 21:295-302(1980).
RN [3]
RP SEQUENCE OF 1-68.
RX MEDLINE=82242310; PubMed=6284987;
RA Dorner A.J., Dorner L.F., Larsen G.R., Wimmer E., Anderson C.W.;
RT "Identification of the initiation site of poliovirus polyprotein
RT synthesis.";
RL J. Virol. 42:1017-1028(1982).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.
RX MEDLINE=85300512; PubMed=2594218;
RA Hogle J.M., Chow M., Filman D.J.;
RT "Three-dimensional structure of poliovirus at 2.9-A resolution.";
RL Science 229:1358-1365(1985).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.
RX MEDLINE=95120467; PubMed=7820548;
RA Grant R.A., Hirenmath C.N., Filman D.J., Syed R., Andries K.,
RA Hogle J.M.;
RT "Structures of poliovirus complexes with anti-viral drugs:
RT implications for viral stability and drug design.";
RL Curr. Biol. 4:784-797(1994).
CC -I- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -I- CATALYTIC ACTIVITY: Selective cleavage of Gln-[Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Glu may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -I- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC -I- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -I- PTM: VP3 IS LINKED, PROBABLY BY TYR-1544, TO THE URIDYLATE RESIDUE
CC AT THE 5' END OF THE GENOME RNA. IT MAY PLAY A ROLE IN THE
CC INITIATION OF RNA SYNTHESIS AND IT MAY ALSO BE INVOLVED IN
CC MORPHOGENESIS.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -I- CAUTION: SEE ALSO VERSION 2 OF THIS PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V01148; CAA24446.1; -.
CC -----
```

DR PIR; A93258; GNNYIP.  
DR PDB; 2PLV; 15-JUL-93.  
DR PDB; IFPT; 31-MAR-95.  
DR PDB; IPOV; 07-DEC-95.  
DR PDB; 1VBD; 11-JUL-96.  
DR PDB; IAL2; 19-NOV-97.  
DR PDB; IAR6; 03-DEC-97.  
DR PDB; IAR7; 03-DEC-97.  
DR PDB; IAR8; 03-DEC-97.  
DR PDB; IAR9; 03-DEC-97.  
DR PDB; IASJ; 03-DEC-97.  
DR PDB; IPO1; 03-DEC-97.  
DR PDB; IPO2; 03-DEC-97.  
DR MEROPS; C03.001; -.  
DR MEROPS; C03.020; -.  
DR InterPro; IPR003593; AAA ATPase.  
DR InterPro; IPR004004; Calici\_pol\_hel.  
DR InterPro; IPR009003; Cys\_Ser\_rypsein.  
DR InterPro; IPR000199; Pept\_3C\_picorn.  
DR InterPro; IPR000081; Peptidase\_C3.  
DR InterPro; IPR003138; Pico\_P1A.  
DR InterPro; IPR002527; Pico\_P2B.  
DR InterPro; IPR001676; RNv.  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_P5vir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00548; Cys-protease-3C; 1.  
DR Pfam; PF02226; Pico\_P1A; 1.  
DR Pfam; PF00947; Pico\_P2A; 1.  
DR Pfam; PF01552; Pico\_P2B; 1.  
DR Pfam; PF00073; rhv; 3.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUSNS.  
DR ProDom; PD001125; Cys\_protease\_3C; 1.  
DR ProDom; PD001306; Pico\_P2A; 1.  
DR ProDom; PD001274; Pico\_P2B; 1.  
DR SMART; SM00382; AAA; 1.  
KW Polypeptide; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;  
KW 3D-structure; Lipoprotein.  
FT INIT\_MET 0 0  
FT CHAIN 1 68 COAT PROTEIN VP4.  
FT CHAIN 69 339 COAT PROTEIN VP2.  
FT CHAIN 340 577 COAT PROTEIN VP3.  
FT CHAIN 578 879 COAT PROTEIN VP1.  
FT CHAIN 880 1028 CORE PROTEIN P2A.  
FT CHAIN 1029 1125 CORE PROTEIN P2B.  
FT CHAIN 1126 1454 CORE PROTEIN P2C.  
FT CHAIN 1455 1541 CORE PROTEIN P3A.  
FT CHAIN 1542 1563 GENOME-LINKED PROTEIN VPG.  
FT CHAIN 1564 1745 PICORNAIN 3C.  
FT CHAIN 1746 2206 RNA-DIRECTED RNA POLYMERASE P3D.  
FT LIPID 1 1 N-myristoyl glycine (in host).  
FT BINDING 1544 1544 URIDYLATE AT THE 5' END OF THE GENOME RNA.  
FT ACT\_SITE 1710 1710 PROTEASE (POTENTIAL).  
FT STRAND 3 6  
FT STRAND 25 28  
FT HELIX 35 37  
FT STRAND 45 45  
FT TURN 49 49  
FT HELIX 50 53  
FT STRAND 56 56  
FT TURN 62 63  
FT TURN 74 75  
FT TURN 79 80  
FT STRAND 82 86  
FT TURN 87 88  
FT STRAND 89 93  
FT STRAND 96 101  
FT HELIX 102 104

FT TURN 114  
FT STRAND 122  
FT HELIX 125  
FT TURN 128  
FT STRAND 132  
FT STRAND 137  
FT TURN 141  
FT STRAND 146  
FT TURN 151  
FT HELIX 152  
FT TURN 155  
FT HELIX 158  
FT STRAND 167  
FT TURN 184  
FT STRAND 186  
FT TURN 197  
FT STRAND 202  
FT HELIX 213  
FT STRAND 219  
FT STRAND 223  
FT TURN 226  
FT TURN 235  
FT STRAND 242  
FT STRAND 245  
FT HELIX 246  
FT TURN 249  
FT HELIX 255  
FT STRAND 262  
FT TURN 267  
FT STRAND 272  
FT STRAND 286  
FT TURN 288  
FT STRAND 291  
FT STRAND 294  
FT STRAND 313  
FT STRAND 336  
FT TURN 347  
FT TURN 354  
FT STRAND 362  
FT TURN 365  
FT STRAND 378  
FT STRAND 381  
FT HELIX 382  
FT TURN 387  
FT STRAND 390  
FT STRAND 396  
FT TURN 398

Query Match 54.3%; Score 19; DB 1; Length 2206;

Best Local Similarity 30.8%; Pred. No. 1.2e+03;

Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16

Db 1342 YVLASTNSSRISP 1354

RESULT 4

POLH\_POLIM

ID POLH\_POLIM STANDARD; PRT; 2208 AA.

AC P03300; Q84879; Q84880; Q89679;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein (Contains: Coat protein VP4 (P1A); Coat protein VP2

DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A

DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein

DE P3A; Genome-linked protein VPG (P3B); Picornain 3C (EC 3.4.22.28)

DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)]

DE (Version 2).

OS Poliovirus type 1 (strain Mahoney).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Enterovirus.

OX NCBI\_TaxID=12081;  
RN [1] SEQUENCE FROM N.A.  
RX MEDLINE=82060159; PubMed=6272282;  
RA Racaniello V.R., Baltimore D.;  
RT "Molecular cloning of poliovirus cDNA and determination of the  
RL complete nucleotide sequence of the viral genome.";  
RN Proc. Natl. Acad. Sci. U.S.A. 78:4887-4891(1981).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.  
RX MEDLINE=85300512; PubMed=2994218;  
RA Hogle J.M., Chow M., Filman D.J.;  
RT "Three-dimensional structure of poliovirus at 2.9-A resolution.";  
RN Science 229:1358-1365(1985).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.  
RX MEDLINE=95120467; PubMed=7820548;  
RA Grant R.A., Hiramath C.N., Filman D.J., Syed R., Andries K.,  
RH Hogle J.M.;  
RT "Structures of poliovirus complexes with anti-viral drugs:  
RL implications for viral stability and drug design.";  
RN Curr. Biol. 4:784-797(1994).  
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN  
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.  
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the  
CC poliovirus polyprotein. In other picornavirus reactions Glu may be  
CC substituted for Gln, and Ser or Thr for Gly.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA) (N).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -!- CAUTION: SEE ALSO VERSION 1 OF THIS PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; V01149; CAA24461.1; -;  
DR PIR; A03898; GNYP2P.  
DR PDB; 2PLV; 15-JUL-93.  
DR PDB; 1FPT; 31-MAR-95.  
DR PDB; 1POV; 07-DEC-95.  
DR PDB; 1VBD; 11-JUL-96.  
DR PDB; 1AL2; 19-NOV-97.  
DR PDB; 1AR6; 03-DEC-97.  
DR PDB; 1AR7; 03-DEC-97.  
DR PDB; 1AR8; 03-DEC-97.  
DR PDB; 1AR9; 03-DEC-97.  
DR PDB; 1ASJ; 03-DEC-97.  
DR PDB; 1P01; 03-DEC-97.  
DR PDB; 1P02; 03-DEC-97.  
DR PDB; 1LIN; 10-APR-02.  
DR MEROPS; C03.001; -;  
DR MEROPS; C03.020; -;  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR004004; Calici\_pol\_hel.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR000199; Pept\_3C\_picorn.  
DR InterPro; IPR000081; Peptidase\_C3.  
DR InterPro; IPR003138; Pico\_P1A.  
DR InterPro; IPR002527; Pico\_P2B.  
DR InterPro; IPR001676; Rhv.  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_Psvir.

DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00548; Cys-protease-3C; 1.  
DR PDB; 2O2226; Pico\_P1A; 1.  
DR Pfam; PF00947; Pico\_P2A; 1.  
DR PDB; PF01552; Pico\_P2B; 1.  
DR Pfam; PF00073; rhv; 3.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICVIRUSNS.  
DR ProDom; PD001125; Cys\_protease\_3C; 1.  
DR ProDom; PD001306; Pico\_P2A; 1.  
DR ProDom; PD001274; Pico\_P2B; 1.  
DR SMART; SM00382; AAA; 1.  
KW Polyprotein; Coat protein; Core protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;  
KW 3D-structure; Lipoprotein.  
FT INIT\_MET 0  
FT CHAIN 1 68 COAT PROTEIN VP4.  
FT CHAIN 69 340 COAT PROTEIN VP2.  
FT CHAIN 341 578 COAT PROTEIN VP3.  
FT CHAIN 579 880 COAT PROTEIN VP1.  
FT CHAIN 881 1029 CORE PROTEIN P2A.  
FT CHAIN 1030 1126 CORE PROTEIN P2B.  
FT CHAIN 1127 1455 CORE PROTEIN P2C.  
FT CHAIN 1456 1542 CORE PROTEIN P3A.  
FT CHAIN 1543 1564 CORE PROTEIN P3B.  
FT CHAIN 1565 1746 GENOME-LINKED PROTEIN VPG.  
FT CHAIN 1747 2208 PICORNAIN 3C.  
FT LIPID 1 1 RNA-DIRECTED RNA POLYMERASE P3D.  
FT ACT\_SITE 1711 1711 N-myristoyl glycine (in host).  
FT ACT\_SITE 1725 1725 PROTEASE (POTENTIAL).  
FT STRAND 3 6 PROTEASE (POTENTIAL).  
FT STRAND 25 28  
FT STRAND 35 37  
FT HELIX 45 45  
FT STRAND 49 49  
FT TURN 50 53  
FT STRAND 56 56  
FT TURN 62 63  
FT TURN 74 75  
FT TURN 79 80  
FT STRAND 82 86  
FT TURN 87 88  
FT STRAND 89 93  
FT STRAND 96 101  
FT HELIX 102 104  
FT TURN 112 114  
FT STRAND 122 122  
FT HELIX 125 127  
FT TURN 128 128  
FT STRAND 132 133  
FT STRAND 137 140  
FT TURN 141 142  
FT STRAND 146 150  
FT TURN 151 151  
FT HELIX 152 154  
FT TURN 155 156  
FT HELIX 158 166  
FT STRAND 167 179  
FT TURN 184 185  
FT STRAND 186 196  
FT TURN 197 197  
FT STRAND 202 202  
FT HELIX 213 216  
FT STRAND 219 221  
FT STRAND 223 224  
FT TURN 226 226  
FT TURN 235 236  
FT STRAND 242 242  
FT STRAND 245 245  
FT HELIX 246 248  
FT TURN 249 252  
FT HELIX 255 260

```

FT STRAND 263 267
FT TURN 268 270
FT STRAND 273 278
FT STRAND 287 291
FT TURN 289 291
FT STRAND 292 292
FT STRAND 295 307
FT STRAND 314 331
FT STRAND 337 337
FT TURN 348 351
FT TURN 355 356
FT TURN 363 363
FT TURN 366 367
FT STRAND 379 380
FT STRAND 382 382
FT STRAND 383 387
FT TURN 388 388
FT STRAND 391 392
FT STRAND 397 397
FT TURN 399 403
FT TURN 405 408
FT STRAND 410 413
FT STRAND 423 426
FT STRAND 429 431
FT TURN 433 437
FT TURN 439 444
FT TURN 445 446
FT STRAND 447 451
FT STRAND 454 460
FT TURN 464 465
FT STRAND 467 467
FT STRAND 469 475
FT STRAND 485 489
FT TURN 490 490
FT STRAND 492 497
FT STRAND 503 508
FT STRAND 517 518
FT TURN 523 525
FT STRAND 529 534
FT STRAND 539 539

Query Match 54.3%; Score 19; DB 1; Length 2208;
Best Local Similarity 30.8%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXXXSP 16
Db 1343 YVLASTNSSRISP 1355

RESULT 5
POLG_POL1S STANDARD; PRT: 2209 AA
AC P03301; Q84881; Q84882; Q84883; Q84884; Q84885; Q84886; Q84887;
AC Q84888; Q84889; Q84890;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (Contains: Coat protein VP4 (PIA); Coat protein VP2
DE (PIB); Coat protein VP3 (PIC); Coat protein VP1 (PID); Picornain 2A
DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
DE P3A; Genome-linked protein VPB (P3B); Picornain 3C (EC 3.4.22.28)
DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
OS Poliovirus type 1 (strain Sabin).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=12082;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83299876; PubMed=6310545;
RA Noncto A., Onata T., Toyoda H., Kuge S., Horie H., Kataoka Y.,
RA Genba Y., Nakano Y., Imura N.;
RT "Complete nucleotide sequence of the attenuated poliovirus Sabin 1

```

```

RT strain genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:5793-5797(1982).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Glu may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- MISCELLANEOUS: THIS VIRUS IS A LIVE VACCINE STRAIN DERIVED FROM
CC THE MAHONEY STRAIN BY SPONTANEOUS MUTATIONS DURING THE ATTENUATION
CC PROCESS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V01150; CAA24465.1; -.
CC MEROPS; C03.001; -.
CC InterPro; IPR004004; Calici_pol_hel.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001199; Pept_3C_picorn.
CC InterPro; IPR000081; Peptidase_C3.
CC InterPro; IPR003138; Pico_P1A.
CC InterPro; IPR002527; Pico_P2B.
CC InterPro; IPR001676; RNV.
CC InterPro; IPR000605; RNA_helicase.
CC InterPro; IPR007095; RNA_pol_DS_PS.
CC InterPro; IPR001205; RNA_pol_P3D.
CC InterPro; IPR007094; RNA_pol_PSVir.
CC InterPro; IPR008975; Viral_cap_coat.
CC Pfam; PF00548; Cys-protease-3C; 1.
CC Pfam; PF02226; Pico_P1A; 1.
CC Pfam; PF00947; Pico_P2A; 1.
CC Pfam; PF01552; Pico_P2B; 1.
CC Pfam; PF00073; rhv_3.
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC PRINTS; PR00918; CALICVIRUSNS.
CC ProDom; PD001125; Cys_protease_3C; 1.
CC ProDom; PD001306; Pico_P2A; 1.
CC ProDom; PD001274; Pico_P2B; 1.
CC Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
CC Lipoprotein.
FT CHAIN 2 69 COAT PROTEIN VP4.
FT CHAIN 70 341 COAT PROTEIN VP2.
FT CHAIN 342 579 COAT PROTEIN VP3.
FT CHAIN 580 881 COAT PROTEIN VP1.
FT CHAIN 882 1030 CORE PROTEIN P2A.
FT CHAIN 1031 1127 CORE PROTEIN P2B.
FT CHAIN 1128 1456 CORE PROTEIN P2C.
FT CHAIN 1457 1543 CORE PROTEIN P3A.
FT CHAIN 1544 1565 GENOME-LINKED PROTEIN VPG.
FT CHAIN 1566 1747 PICORNAIN 3C.
FT CHAIN 1748 2209 RNA-DIRECTED RNA POLYMERASE P3D.
FT LIPID 2 2 N-myristoyl glycine (in host).
FT ACT SITE 1712 1712 PROTEASE (POTENTIAL).
FT ACT SITE 1726 1726 PROTEASE (POTENTIAL).
SQ SEQUENCE 2209 AA; 246576 MW; 9EC1EF4D174A28A4 CRC64;

Query Match 54.3%; Score 19; DB 1; Length 2209;
Best Local Similarity 30.8%; Pred. No. 1.2e+03;

```

```

Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 4 YXXXXXXXRXSP 16
Db 1344 YVLASTNSGRISP 1356
RESULT 6
ID ULA7 HCWVA STANDARD; PRT; 150 AA.
AC P1628;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Hypothetical protein UL107.
GN UL107.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,
RA Freddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17403; CAA35343.1; -.
DR HAMAP; MF_01810; -.
DR InterPro; IPR001708; 60kDa_innermemb.
DR Pfam; PF02096; 60KD_IMP; 1.
KW Transmembrane 12 34 POTENTIAL.
FT TRANSMEM 38 60 POTENTIAL.
FT TRANSMEM 104 126 POTENTIAL.
FT TRANSMEM 192 214 POTENTIAL.
FT TRANSMEM 235 257 POTENTIAL.
SQ SEQUENCE 150 AA; 17373 MW; 042707546C7EB878 CRC64;
Query Match 51.4%; Score 18; DB 1; Length 150;
Best Local Similarity 30.8%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 4 YXXXXXXXRXSP 16
Db 4 YSITVTYDHRTPSP 16
RESULT 7
ID OXAA MYCPA STANDARD; PRT; 353 AA.
AC O9L7M1;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Membrane protein oxaA.
GN OXAA.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RX Zhang Q., Kapur V.;
RT "Genomic organization of the Mycobacterium avium subsp.
RT paratuberculosis origin of replication region."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Required for the insertion of integral membrane proteins

```

```

into the membrane. Probably plays an essential role in the
integration of proteins of the respiratory chain complexes.
Involved in integration of membrane proteins that insert
independently and independently of the Sec translocase complex (By
similarity).
-!- SUBUNIT: Specifically interacts with transmembrane segments of
naeant integral membrane proteins during membrane integration (By
similarity).
-!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
-!- SIMILARITY: Belongs to the OXA1/OXA family. Subfamily 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF222789; AAF33698.1; -.
DR HAMAP; MF_01810; -.
DR InterPro; IPR001708; 60kDa_innermemb.
DR Pfam; PF02096; 60KD_IMP; 1.
KW Transmembrane 12 34 POTENTIAL.
FT TRANSMEM 38 60 POTENTIAL.
FT TRANSMEM 104 126 POTENTIAL.
FT TRANSMEM 192 214 POTENTIAL.
FT TRANSMEM 235 257 POTENTIAL.
SQ SEQUENCE 353 AA; 39246 MW; FCBA68ADB523BA90 CRC64;
Query Match 51.4%; Score 18; DB 1; Length 353;
Best Local Similarity 30.8%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 4 YXXXXXXXRXSP 16
Db 209 YFNSRASVARQSP 221
RESULT 8
ID ASSY METAC STANDARD; PRT; 394 AA.
AC Q8TNY5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate
DE ligase).
GN ARG OR MA2142.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.G., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +

```



```
CC diphosphate + L-argininosuccinate.
CC -!- PATHWAY: Arginine biosynthesis; seventh step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the argininosuccinate synthase family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE010900; AAM05540.1; -.
CC HAMAP; MF_00005; -.
CC InterPro; IPR001518; Arginosuc_synth.
CC Pfam; PF00764; Arginosuc_synth; 1.
CC ProDom; PD003544; Arginosuc_synth; 1.
CC TIGRFAMs; TIGR00032; argG; 1.
CC PROSITE; PS00564; ARGININOSUCCIN_SYN_1; 1.
CC PROSITE; PS00565; ARGININOSUCCIN_SYN_2; 1.
CC Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.
CC KW
CC SEQUENCE 394 AA; 43994 MW; 2B995B9C138E7E7F1 CRC64;
CC
CC Query Match 51.4%; Score 18; DB 1; Length 394;
CC Best Local Similarity 30.8%; Pred. No. 3.8e+02;
CC Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
CC
CC Qy 4 YXXXXXXXRXSP 16
CC | | | |
CC Db 341 YKGALTILARSSP 353
CC
CC RESULT 9
CC ASSY_METBA
CC ID ASSY_METBA STANDARD; PRT; 394 AA.
CC AC Q8Q0U5;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Argininosuccinate synthase (EC 6.3.4.5) (Citruilline--aspartate
CC DE ligase).
CC OS ARGG OR MM0037.
CC GN Methanosarcina mazei (Methanosarcina frisia).
CC OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
CC OC Methanosarcinaceae; Methanosarcina.
CC OX NCBI_TaxID=2209;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Goel / Goel / ATCC BAA-199 / DSM 3647 / OCM 88;
CC RX MEDLINE=22120827; PubMed=12125824;
CC RA Deppe-meier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
CC RA Martinez-Arias R., Henne A., Wieser A., Baeumer S., Jacobi C.,
CC RA Brueggemann H., Lienard T., Christmann A., Boencke M., Steckel S.,
CC RA Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
CC RA Fritz H.-J., Gottschalk G.
CC RT "The genome of Methanosarcina mazei: evidence for lateral gene
CC RT transfer between Bacteria and Archaea."
CC RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC CC -!- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +
CC CC diphosphate + L-argininosuccinate.
CC CC -!- PATHWAY: Arginine biosynthesis; seventh step.
CC CC -!- SUBUNIT: Homotetramer (By similarity).
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC CC -!- SIMILARITY: Belongs to the argininosuccinate synthase family.
CC CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE013223; AAM29733.1; ALT_INIT.
CC HAMAP; MF_00005; -.
CC InterPro; IPR001518; Arginosuc_synth.
CC Pfam; PF00764; Arginosuc_synth; 1.
CC ProDom; PD003544; Arginosuc_synth; 1.
CC TIGRFAMs; TIGR00032; argG; 1.
CC PROSITE; PS00564; ARGININOSUCCIN_SYN_1; 1.
CC PROSITE; PS00565; ARGININOSUCCIN_SYN_2; 1.
CC Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.
CC KW
CC SEQUENCE 394 AA; 43988 MW; 2770F2095D058B05 CRC64;
CC
CC Query Match 51.4%; Score 18; DB 1; Length 394;
CC Best Local Similarity 30.8%; Pred. No. 3.8e+02;
CC Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
CC
CC Qy 4 YXXXXXXXRXSP 16
CC | | | |
CC Db 341 YKGALTILARSSP 353
CC
CC RESULT 10
CC ASSY_METBA
CC ID ASSY_METBA STANDARD; PRT; 396 AA.
CC AC P13257;
CC DT 01-JAN-1990 (Rel. 13, Created)
CC DT 01-JAN-1990 (Rel. 13, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Argininosuccinate synthase (EC 6.3.4.5) (Citruilline--aspartate
CC DE ligase).
CC GN ARGG.
CC OS Methanosarcina barkeri.
CC OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
CC OC Methanosarcinaceae; Methanosarcina.
CC OX NCBI_TaxID=2208;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=MS / DSM 800;
CC RX MEDLINE=88257029; PubMed=3133361;
CC RA Morris C.J., Reeve J.N.;
CC RT "Conservation of structure in the human gene encoding
CC RT argininosuccinate synthetase and the argG genes of the archaeobacteria
CC RT Methanosarcina barkeri MS and Methanococcus vannielii."
CC RL J. Bacteriol. 170:3125-3130(1988).
CC CC -!- CATALYTIC ACTIVITY: ATP + L-citrulline + L-aspartate = AMP +
CC CC diphosphate + L-argininosuccinate.
CC CC -!- PATHWAY: Arginine biosynthesis; seventh step.
CC CC -!- SUBUNIT: Homotetramer.
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC CC -!- SIMILARITY: Belongs to the argininosuccinate synthase family.
CC CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M21314; AAA72677.1; -.
CC PIR; B28180; AJMZRB.
CC HAMAP; MF_00005; -.
CC InterPro; IPR001518; Arginosuc_synth.
CC Pfam; PF00764; Arginosuc_synth; 1.
CC ProDom; PD003544; Arginosuc_synth; 1.
CC TIGRFAMs; TIGR00032; argG; 1.
CC PROSITE; PS00564; ARGININOSUCCIN_SYN_1; 1.
CC PROSITE; PS00565; ARGININOSUCCIN_SYN_2; 1.
CC Arginine biosynthesis; Ligase; ATP-binding.
CC KW
```

SQ SEQUENCE 396 AA; 44441 MW; BF9DA4207D1ACC6B CRC64;

Query Match 51.4%; Score 18; DB 1; Length 396;  
 Best Local Similarity 30.8%; Pred. No. 3.8e+02;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16

Db 342 YKALTILARSP 354

RESULT 11

PHYB ASPAW

ID PHYB ASPAW STANDARD; PRT; 479 AA.

AC P34755;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE 3-phytase B precursor (EC 3.1.1.3.8) (Myo-inositol-hexaphosphate

DE 3-phosphohydrolase B) (pH 2.5 optimum acid phosphatase).

GN PHYB OR APH.

OS Aspergillus awamori.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OX NCBI\_TaxID=105351;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ALK0243;

RX MEDLINE=94040796; PubMed=8224894;

RA Piddington C.S., Houston C.S., Paloheimo M.T., Cantrell M.A.,

RA Miettinen-Oinonen A., Nevalainen H., Rambossek J.A.;

RT "The cloning and sequencing of the genes encoding phytase (phy) and

RT pH 2.5-optimum acid phosphatase (aph) from Aspergillus niger var.

RT awamori.";

RL Gene 133:55-62 (1993).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).

RX MEDLINE=9226417; PubMed=10329192;

RA Kostrewa D., Wyss M., D'Arcy A., van Loon A.P.;

RT "Crystal structure of Aspergillus niger pH 2.5 acid phosphatase at

RT 2.4-A resolution.";

RL J. Mol. Biol. 288:965-974 (1999).

CC -!- FUNCTION: Catalyzes the hydrolysis of inorganic orthophosphate

CC from phytate.

CC -!- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = 1D-

CC myo-inositol 1,2,4,5,6-pentakisphosphate + phosphate.

CC -!- SUBUNIT: Homodimer.

CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; L02420; AAA16897.1; --

DR EMBL; JN0890; JN0890.

DR PDB; 1QFX; 19-APR-00.

DR InterPro; IPR000560; HisAc\_phsphtse.

DR Pfam; PF00328; acid phosphat; 1.

DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.

DR PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.

KW Hydrolase; Glycoprotein; Signal; 3D-structure.

FT SIGNAL 1 19 BY SIMILARITY.

FT CHAIN 20 479 3-PHYTASE B.

FT ACT\_SITE 82 82 NUCLEOPHILIC ACCEPTOR.

FT ACT\_SITE 337 337 PROTON DONOR.

FT DISULFID 71 387

FT DISULFID 128 472

FT DISULFID 216 441

FT DISULFID 225 298

FT DISULFID 413 191

FT CARBOHYD 315 315

FT CARBOHYD 458 458

FT HELIX 41 43

FT HELIX 45 47

FT TURN 50 51

FT TURN 69 70

FT STRAND 71 81

FT STRAND 85 85

FT HELIX 89 103

FT HELIX 112 119

FT HELIX 127 129

FT TURN 130 131

FT STRAND 133 133

FT TURN 138 139

FT HELIX 141 156

FT HELIX 157 159

FT STRAND 166 171

FT STRAND 173 173

FT HELIX 174 188

FT HELIX 189 191

FT HELIX 192 195

FT STRAND 196 201

FT HELIX 205 207

FT TURN 208 209

FT TURN 224 227

FT TURN 233 233

FT HELIX 234 245

FT TURN 247 248

FT HELIX 253 266

FT TURN 267 269

FT TURN 274 275

FT HELIX 276 278

FT HELIX 281 298

FT TURN 301 302

FT TURN 304 305

FT HELIX 306 323

FT HELIX 325 328

FT STRAND 331 336

FT HELIX 338 348

FT TURN 349 349

FT TURN 364 365

FT HELIX 370 372

FT TURN 373 373

FT TURN 376 377

FT STRAND 379 387

FT STRAND 390 390

FT TURN 391 392

FT STRAND 393 393

FT STRAND 396 403

FT TURN 404 405

FT STRAND 406 408

FT STRAND 410 410

FT TURN 411 412

FT STRAND 413 413

FT HELIX 417 419

FT STRAND 421 422

FT HELIX 423 433

FT HELIX 437 440

FT TURN 441 442

FT TURN 445 446

FT TURN 453 455

FT TURN 462 463

FT TURN 472 473

FT STRAND 476 476

SQ SEQUENCE 479 AA; 52678 MW; 4F8E0F3778CC3B08 CRC64;

Query Match 51.4%; Score 18; DB 1; Length 479;  
 Best Local Similarity 30.8%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16

Db 457 YNTTTELNYRSP 469

## RESULT 12

PHYB ASPNG STANDARD; PRT; 479 AA.  
 ID - PHYB ASPNG  
 AC P34754;  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 3-phytase B precursor (EC 3.1.3.8) (Myo-inositol-hexakisphosphate  
 DE 3-phosphohydrolase B) (3 phytase B) (Myo-inositol hexakisphosphate  
 DE phosphohydrolase B).  
 GN PHYB.

OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eutriales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5661;

[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 20-101; 133-141 AND 376-399.

RX MEDLINE=93371452; PubMed=7916610;

RA Ehrlich K.C., Montalbano B.G., Mullaney E.J., Dischinger H.C. Jr.,

RA Ullah A.H.J.;

RT Identification and cloning of a second phytase gene (phyB) from

RT Aspergillus niger (ficus).";

RL Biochem. Biophys. Res. Commun. 195:53-57(1993).

CC -!- FUNCTION: Catalyzes the hydrolysis of inorganic orthophosphate

CC from phytate.

CC -!- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = 1D-

CC myo-inositol 1,2,4,5,6-pentakisphosphate + phosphate.

CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; L20567; AAA02934.1; -.

CC HSRP; P34755; IQFX.

CC InterPro; IPR000560; HisAc\_phsphtase.

CC Pfam; PF00328; acid phosphat; 1.

CC PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; 1.

CC PROSITE; PS00778; HIS\_ACID\_PHOSPHAT\_2; 1.

CC Hydrolase; Glycoprotein; Signal.

CC SIGNAL 1 19

FT CHAIN 20 479 3-PHYTASE B.

FT ACT\_SITE 81 81 REQUIRED FOR BINDING SUBSTRATE

FT ACT\_SITE 82 82 (BY SIMILARITY).

FT ACT\_SITE 83 83 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).

FT ACT\_SITE 84 84 PROTON DONOR (BY SIMILARITY).

FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 479 479 395D4DA2B50DFC4 CRC64;

FT SEQUENCE 479 AA; 52611 MW; 395D4DA2B50DFC4 CRC64;

Query Match 51.4%; Score 18; DB 1; Length 479;

Best Local Similarity 30.8%; Pred. No. 4.7e+02;

Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

4 YXXXXXXXRXSP 16

457 YNTTTELNYRSP 469

Qy

Db

## RESULT 13

GYG2 HUMAN

ID - GYG2 HUMAN

AC P15488; O15486; O15487; O15489; O15490;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Glycogenin-2 (EC 2.4.1.186) (GN-2) (GN2).

GN GYG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

[1]

SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC TISSUE=Liver;

RX MEDLINE=98010589; PubMed=9346895;

RA Mu J., Skurat A.V., Roach P.J.;

RT "Glycogenin-2, a novel self-glycosylating protein involved in liver

RT glycogen biosynthesis.";

RL J. Biol. Chem. 272:27589-27597(1997).

[2]

SEQUENCE FROM N.A.

RP MEDLINE=20184741; PubMed=10721716;

RA Zhai L., Mu J., Zong H., Depaoli-Roach A.A., Roach P.J.;

RT "Structure and chromosomal localization of the human glycogenin-2 gene

RT GYG2.";

RL Gene 242:229-235(2000).

[3]

SEQUENCE FROM N.A. (ISOFORM BETA).

RP TISSUE=Skin;

RX MEDLINE=22388257; PubMed=12477932;

RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausberg R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Alschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaez T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Sange C.,

RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[4]

CHARACTERIZATION, AND CARBOHYDRATE-LINKAGE SITE TYR-228.

RP MEDLINE=99074257; PubMed=9857012;

RA Mu J., Roach P.J.;

RT "Characterization of human glycogenin-2, a self-glycosylating

RT initiator of liver glycogen metabolism.";

RL J. Biol. Chem. 273:34850-34856(1998).

[5]

MASS SPECTROMETRY.

RP TISSUE=Breast cancer;

RX MEDLINE=21829512; PubMed=11840567;

RA Harris R.A., Yang A., Stein R.C., Lucy K., Brusten L., Herath A.,

RA Parekh R., Waterfield M.D., O'Hare M.J., Neville M.A., Page M.J.,

RA Zvelebil M.J.;

RT "Cluster analysis of an extensive human breast cancer cell line

RT protein expression map database.";

RL Proteomics 2:212-223(2002).

CC -!- FUNCTION: Self-glycosylates, via an inter-subunit mechanism, to

CC form an oligosaccharide primer that serves as substrate for

CC glycogen synthase.

CC -!- CATALYTIC ACTIVITY: UDP-glucose + glycogenin = UDP +

CC glycosylglycogenin.  
 CC -!- COFACTOR: Self-glycosylation is dependent on the presence of  
 CC divalent metal ions of which manganese ion is the most effective.  
 CC -!- PATHWAY: Glycogen biosynthesis.  
 CC -!- SUBUNIT: Homodimer, tightly complexed to glycogen synthase.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=6;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=Alpha;  
 CC IsoId=O15488-1; Sequence=Displayed;  
 CC Name=Beta;  
 CC IsoId=O15488-2; Sequence=VSP\_001770;  
 CC Name=Gamma;  
 CC IsoId=O15488-3; Sequence=VSP\_001771;  
 CC Name=Delta;  
 CC IsoId=O15488-4; Sequence=VSP\_001772;  
 CC Name=Epsilon;  
 CC IsoId=O15488-5; Sequence=VSP\_001773;  
 CC Name=Zeta;  
 CC IsoId=O15488-6; Sequence=VSP\_001774;  
 CC -!- TISSUE SPECIFICITY: Expressed preferentially in liver, heart, and  
 CC pancreas.  
 CC -!- PTM: SELF-GLYCOSYLATED BY THE TRANSFER OF GLUCOSE RESIDUES FROM  
 CC UDP-GLUCOSE TO ITSELF, FORMING AN ALPHA-1,4-GLYCAN OF AROUND 10  
 CC RESIDUES ATTACHED TO TYR-228.  
 CC -!- MASS SPECTROMETRY: MW=55211.89; METHOD=MALDI.  
 CC -!- SIMILARITY: Belongs to the glycogenin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U94362; AAB84377.1; -  
 DR EMBL; U94363; AAB84378.1; -  
 DR EMBL; U94364; AAB84379.1; -  
 DR EMBL; U94357; AAB84373.1; -  
 DR EMBL; U94358; AAB84374.1; -  
 DR EMBL; U94360; AAB84375.1; -  
 DR EMBL; U94361; AAB84376.1; -  
 DR EMBL; AF179624; AAF61855.1; -  
 DR EMBL; AF179615; AAF61855.1; JOINED.  
 DR EMBL; AF179616; AAF61855.1; JOINED.  
 DR EMBL; AF179617; AAF61855.1; JOINED.  
 DR EMBL; AF179618; AAF61855.1; JOINED.  
 DR EMBL; AF179619; AAF61855.1; JOINED.  
 DR EMBL; AF179620; AAF61855.1; JOINED.  
 DR EMBL; AF179621; AAF61855.1; JOINED.  
 DR EMBL; AF179622; AAF61855.1; JOINED.  
 DR EMBL; AF179623; AAF61855.1; JOINED.  
 DR EMBL; BC023152; AAB23152.1; -  
 DR GK; O15488; -  
 DR MIM; 300198; -  
 DR GO; GO:0005625; C-soluble fraction; TAS.  
 DR GO; GO:0008466; F-glycogenin glucosyltransferase activity; TAS.  
 DR GO; GO:0005978; P-glycogenin biosynthesis; TAS.  
 DR InterPro; IPR002495; Glyco trans 8.  
 DR Pfam; PF01501; Glyco trans 8; 1.  
 DR Transfaser; Glycogen biosynthesis; Alternative splicing;  
 KW Glycoprotein; Polymorphism.  
 FT ACT SITE 119 119 BY SIMILARITY.  
 FT CARBOHYD 228 228 O-LINKED (GLC...).  
 FT VARSPPLIC 3 33 Missing (in isoform Beta).  
 FT FT FT  
 FT VARSPPLIC 3 42 Missing (in isoform Gamma).  
 FT FT FT  
 FT VARSPPLIC 378 448 Missing (in isoform Delta).  
 FT FT FT  
 FT VARSPPLIC 407 501 Missing (in isoform Epsilon).  
 FT FT FT

FT FT VARSPPLIC 413 448 /FTid=VSP\_001773.  
 FT Missing (in isoform Zeta).  
 FT /FTid=VSP\_001774.  
 FT V -> A.  
 FT /FTid=VAR\_010401.  
 FT MUTAGEN 228 228 Y->F: LOSS OF ACTIVITY.  
 FT MUTAGEN 230 230 Y->F: NO LOSS OF ACTIVITY.  
 FT CONFLICT 313 313 H -> R (IN REF. 3).  
 FT CONFLICT 413 413 MISSING (IN REF. 1; AAB84378).  
 FT CONFLICT 462 464 EKV -> AGI (IN REF. 1; AAB84376).  
 SQ SEQUENCE 501 AA; 55211 MW; 2EDE05FDAD5A7657 CRC64;  
 Query Match 51.4%; Score 18; DB 1; Length 501;  
 Best Local Similarity 30.8%; Pred. No. 4.9e+02;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 4 YXXXXXXXXXXSP 16  
 DB 295 YKSVQAGEARASP 307  
 RESULT 14  
 NRGI\_CHICK  
 ID NRGI\_CHICK STANDARD; PRT; 602 AA.  
 AC Q05199; O73750; O73751; O73752;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Pro-neuregulin-1 precursor (Pro-NRG1) [Contains: Neuregulin-1  
 (Acetylcholine receptor inducing activity) (ARIA)].  
 GN NRGI OR ARIA.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.  
 RC STRAIN=White leghorn; TISSUE=Brain;  
 RX MEDLINE=93201602; PubMed=8453670;  
 RA Falls D.L., Rosen K.M., Corfas G., Lane W.S., Fischbach G.D.;  
 RT "ARIA, a protein that stimulates acetylcholine receptor synthesis, is  
 RT a member of the neu ligand family.";  
 RL Cell 72:801-815(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4).  
 RC TISSUE=Brain, and Spinal cord;  
 RX MEDLINE=98150951; PubMed=9491987;  
 RA Yang X., Kuo Y., Devay P., Yu C., Role L.;  
 RT "A cysteine-rich isoform of neuregulin controls the level of  
 RT expression of neuronal nicotinic receptor channels during  
 RT synaptogenesis.";  
 RL Neuron 20:255-270(1998).  
 CC -!- FUNCTION: Direct ligand for the ERBB tyrosine kinase receptors.  
 CC The multiple isoforms perform diverse functions: Cysteine-rich  
 CC domain containing isoforms (isoforms 2-4) probably regulate the  
 CC expression of nicotinic acetylcholine receptors at developing  
 CC interneuronal synapses. The IG-NRG isoform is required for the  
 CC initial induction and/or maintenance of the mature levels of  
 CC acetylcholine receptors at neuromuscular synapses.  
 CC -!- SUBCELLULAR LOCATION: Exists as a type I membrane protein and as a  
 CC proteolytically released soluble growth factor form. The membrane-  
 CC bound form does not seem to be active (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1; Synonyms=ARIA, IG-NRG;  
 CC IsoId=Q05199-1; Sequence=Displayed;  
 CC Name=2; Synonyms=CRD-NRG-BETA1A;  
 CC IsoId=Q05199-2; Sequence=VSP\_003445;  
 CC Note=The EGF-like domain is replaced by a Cysteine-rich domain  
 CC (CRD);

CC Name=3; Synonyms=CRD-NRG-BETA2A;  
 CC IsoId=Q05199-3; Sequence=VSP\_003445, VSP\_003446;  
 CC Note-The EGF-like domain is replaced by a Cysteine-rich domain  
 CC (CRD);  
 CC Name=4; Synonyms=CRD-NRG-BETA2B;  
 CC IsoId=Q05199-4; Sequence=VSP\_003445, VSP\_003446, VSP\_003447,  
 CC Note-The EGF-like domain is replaced by a Cysteine-rich domain  
 CC (CRD);  
 CC -!- DEVELOPMENTAL STAGE: Isoforms 2-4 are detected at embryonic day 4  
 CC (ED4) in both visceral and somatic motor neurons of spinal cord  
 CC and is highest at ED6. Isoform 1 is not expressed until ED 6 in  
 CC spinal cord. At ED 11 both isoforms display comparable levels.  
 CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation  
 CC of trafficking and proteolytic processing. Regulation of the  
 CC proteolytic processing involves initial intracellular domain  
 CC dimerization (By similarity).  
 CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like  
 CC domain.  
 CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the  
 CC external face leads to the release of the soluble growth factor  
 CC form.  
 CC -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By  
 CC similarity).  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -!- SIMILARITY: Belongs to the neurogulin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L11264; AAA49037.1; -;  
 DR EMBL; AF045654; AAC05670.1; -;  
 DR EMBL; AF045655; AAC05671.1; -;  
 DR EMBL; AF045656; AAC05672.1; -;  
 DR PIR; A45769; A45769.  
 DR HSSP; Q12784; 1HRE.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR002154; Neurogulin.  
 DR Pfam; PF00047; ig; 1.  
 DR Pfam; PF02158; Neurogulin; 1.  
 DR PRINTS; PR01089; NEUREGULIN.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00408; IGC2; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
 DR PROSITE; PS50026; EGF\_3; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;  
 KW Transmembrane; Alternative splicing.  
 FT CHAIN 1 602  
 FT CHAIN 1 205  
 FT CHAIN 1 206  
 FT DOMAIN 1 206  
 FT TRANSMEM 1 206  
 FT DOMAIN 207 229  
 FT DOMAIN 230 602  
 FT DOMAIN 29 123  
 FT DOMAIN 125 136  
 FT DOMAIN 137 181  
 FT DOMAIN 137 181  
 FT DISULFID 49 105  
 FT DISULFID 141 155  
 FT DISULFID 149 169  
 FT DISULFID 171 180  
 FT CARBOHYD 21 21  
 FT CARBOHYD 113 113  
 FT CARBOHYD 126 126

FT VARSPLIC 1 127  
 FT MWATSEGLQVSLAPTQTDVNSVNTVPKLEKMKNQEVAV  
 FT GOKLVRCETTSYEPALRFKMLKNGKITKQNRPNVPIPK  
 FT KKKYSELHIVRATLADAGEYACVSKSLGNDSTKASVIT  
 FT DTNA -> MSEVGTFFPSQAQLSPDASLGLPABENMPG  
 FT PHREDSRVGVAGLASTCCVCLAEARLKGCLNSEKICIAPI  
 FT LACLLSLCLTAGLKWVFDKIFEDYDSPTHLDPGRIGQDPR  
 FT STVDPTALUSAWPSEVIAVSPFIPSPLESKAEVTVQDSSLV  
 FT PSRPFQPSLVNRIIDVGLMSSATSLSPSLEPTTASQAQ  
 FT ATEINLOTAPKLS (in isoform 2, isoform 3  
 FT and isoform 4).  
 FT /FTid=VSP\_003445.  
 FT Missing (in isoform 3 and isoform 4).  
 FT /FTid=VSP\_003446.  
 FT VSAAMTPARMSPVDFHTP -> HTPPTSLLLAGKVSRLVS  
 FT (in isoform 4).  
 FT /FTid=VSP\_003447.  
 FT Missing (in isoform 4).  
 FT /FTid=VSP\_003448.  
 FT SQ SEQUENCE 602 AA; 67453 MW; 4183C0E56CED346 CRC64;  
 Query Match 51.4%; Score 18; DB 1; Length 602;  
 Best Local Similarity 30.8%; Pred. No. 6e+02;  
 Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 4 YXXXXXXXRXSP 16  
 DB 387 YVSAMTTPARMSP 399  
 RESULT 15  
 ID\_NRG1\_HUMAN STANDARD; PRT; 639 AA.  
 AC Q02297; Q14667; P98202; Q02299; Q07110; Q12779;  
 AC Q12780; Q12781; Q12782; Q12783; Q12784; Q9UPE3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Pro-neuregulin-1 precursor (Pro-NRG1) [Contains: Neuregulin-1 (Neu  
 DE differentiation factor) (Heregulin) (HRG) (Breast cancer cell  
 DE differentiation factor p45) (Acetylcholine receptor inducing activity)  
 DE (ARIA) (Sensory and motor neuron-derived factor) (Glial growth  
 DE factor)].  
 GN NRG1 OR HGL OR NDF OR HRGA OR GGF OR SMDF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 6; 7 AND 8), AND PARTIAL  
 RP SEQUENCE.  
 RX MEDLINE=92271253; PubMed=1350381;  
 RA Holmes W.B., Sliwkowski M.X., Akita R.W., Henzel W.J., Lee J.,  
 RA Park J.W., Yansura D., Abadi N., Raab H., Lewis G.D., Shepard H.M.,  
 RA Kuang W.-J., Wood W.I., Goeddel D.V., Vanden R.L.;  
 RT "Identification of heregulin, a specific activator of p185erbB2.";  
 RL Science 256:1205-1210(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 6; 7 AND 8).  
 RC TISSUE=Kidney adenocarcinoma, and Pituitary;  
 RX MEDLINE=94158663; PubMed=7509448;  
 RA Wen D., Suggs S.V., Karunagaran D., Liu N., Cupples R.L., Luo Y.,  
 RA Janssen A.M., Ben-Baruch N., Trollinger D.B., Jacobsen V.L.,  
 RA Meng S.-Y., Lu H.S., Hu S., Chang D., Yang W., Yanigahara D.,  
 RA Koski R.A., Yarden Y.;  
 RT "Structural and functional aspects of the multiplicity of Neu  
 RT differentiation factors.";  
 RL Mol. Cell. Biol. 14:1909-1919(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=92208945; PubMed=1348215;  
 RA Peles E., Bacus S.S., Koski R.A., Lu H.S., Wen D., Ogden S.G.,  
 RA Levy R.B., Yarden Y.;  
 RT "Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein

that induces differentiation of mammary tumor cells.";

Cell 69:205-216(1992).

[4]

SEQUENCE FROM N.A. (ISOFORMS 8 AND 9).

TISSUE=Brain;

MEDLINE=93205115; PubMed=8096067;

Marchionni M.A., Goodearl A.D.J., Chen M.S., Birmingham-McDonogh O., Kirk C., Hendricks M., Danehy F., Misumi D., Sudhalter J., Kobayashi K., Wroblewski D., Lynch C., Baldaarre M., Hiles I., Davis J.B., Huan J.J., Totty N.F., Otsu M., McBurney R.N., Waterfield M.D., Stroobant P., Gwynne D.;

"Glial growth factors are alternatively spliced erbB2 ligands expressed in the nervous system.";

Nature 362:312-318(1993).

[5]

SEQUENCE FROM N.A. OF GAMMA-HEREGULIN FUSION PROTEIN.

TISSUE=Breast cancer;

MEDLINE=97472144; PubMed=9333014;

Schaefer G., Fitzpatrick V.D., Sliwkowski M.X.;

"Gamma-hergulin: a novel heregulin isoform that is an autocrine growth factor for the human breast cancer cell line, MDA-MB-175.";

Oncogene 15:1385-1394(1997).

[6]

SEQUENCE OF 1-210 FROM N.A.

Schoumacher F., Herzer S., Flury N., Kueng W., Mueller H., Eppenberger U.;

Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.

[7]

SEQUENCE OF 19-27.

MEDLINE=93366731; PubMed=7699552;

Culoscou J.-M., Plozman G.D., Carlton G.W., Green J.M., Shoyab M.;

"Characterization of a breast cancer cell differentiation factor that specifically activates the HER4/p180erbB4 receptor.";

J. Biol. Chem. 268:18407-18410(1993).

[8]

CHROMOSOMAL TRANSLOCATION.

MEDLINE=99455251; PubMed=10523851;

Wang X.-Z., Jolicoeur M., Conte N., Chaffanet M., Zhang Y., Mozziconacci M.-J., Feiner H., Birnbaum D., Pebusque M.-J., Ron D.;

"Gamma-hergulin is the product of a chromosomal translocation fusing the DOC4 and HGL/NRG1 genes in the MDA-MB-175 breast cancer cell line.";

Oncogene 18:5718-5721(1999).

[9]

CHROMOSOMAL TRANSLOCATION.

MEDLINE=20065180; PubMed=10597312;

Liu X., Baker E., Eyre H.J., Sutherland G.R., Zhou M.;

"Gamma-hergulin: a fusion gene of DOC-4 and neuregulin-1 derived from a chromosome translocation.";

Oncogene 18:7110-7114(1999).

[10]

STRUCTURE BY NMR OF 175-241 (ISOFORM 1).

MEDLINE=94341264; PubMed=8062828;

Nagata K., Kohda D., Hatanaka H., Ichikawa S., Matsuda S., Yamamoto T., Suzuki A., Inagaki F.;

"Solution structure of the epidermal growth factor-like domain of heregulin-alpha, a ligand for p180erbB-4.";

EMBO J. 13:3517-3523(1994).

CC -1- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors, resulting in ligand-stimulated tyrosine phosphorylation and activation of the ERBB receptors. The multiple isoforms perform diverse functions such as inducing growth and differentiation of epithelial, glial, neuronal, and skeletal muscle cells; inducing expression of acetylcholine receptor in synaptic vesicles during the formation of the neuromuscular junction; stimulating lobuloalveolar budding and milk production in the mammary gland and inducing differentiation of mammary tumor cells; stimulating Schwann cell proliferation; implication in the development of the myocardium such as trabeculation of the developing heart.

CC -1- SUBUNIT: The cytoplasmic domain interacts with the LIM domain region of LIMK1 (by similarity).

CC -1- SUBCELLULAR LOCATION: Exists as an type I membrane protein and as

CC

a proteolytically released soluble growth factor form. The membrane-bound form does not seem to be active. The secreted isoform 9 has a signal peptide. The isoform 8 may be nuclear.

CC

-1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=9;

CC Comment=Additional isoforms seem to exist. Isoforms have been classified as type I NRGs (isoforms with an Ig domain and a glycosylation domain, isoforms 1-8), type II NRGs (isoforms with an Ig domain but no glycosylation domain, isoform 9) and type III NRGs (isoforms with a Cys-rich domain, isoform 10). All these isoforms perform distinct tissue-specific functions;

CC Name=1; Synonyms=Alpha;

CC IsoId=Q02297-1; Sequence=Displayed;

CC Name=2; Synonyms=Alpha1A;

CC IsoId=Q02297-2; Sequence=VSP\_003431;

CC Name=3; Synonyms=Alpha2B;

CC IsoId=Q02297-3; Sequence=VSP\_003434, VSP\_003435;

CC Name=4; Synonyms=Alpha3;

CC IsoId=Q02297-4; Sequence=VSP\_003432, VSP\_003433;

CC Name=6; Synonyms=Betal, Beta1A;

CC IsoId=Q02297-6; Sequence=VSP\_003428;

CC Name=7; Synonyms=Beta2;

CC IsoId=Q02297-7; Sequence=VSP\_003427;

CC Name=8; Synonyms=Beta3, GGFHFB1;

CC IsoId=Q02297-8; Sequence=VSP\_003429, VSP\_003430;

CC Name=9; Synonyms=GGF2, GGFHFP2;

CC IsoId=Q02297-9; Sequence=VSP\_003425, VSP\_003426, VSP\_003429, VSP\_003430;

CC Name=10; Synonyms=SMDF;

CC IsoId=Q15491-1; Sequence=External;

CC -1- TISSUE SPECIFICITY: Type I isoforms are the predominant forms expressed in the endocardium. Isoform alpha is expressed in breast, ovary, testis, prostate, heart, skeletal muscle, lung, placenta liver, kidney, salivary gland, small intestine and brain, but not in uterus, stomach, pancreas, and spleen. Isoform 3 is the predominant form in mesenchymal cells and in nonneural organs, whereas isoform 5 is the major neuronal form. Isoform 8 is expressed in spinal cord and brain. Isoform 9 is the major form in skeletal muscle cells; in the nervous system it is expressed in spinal cord and brain. Also detected in adult heart, placenta, lung, liver, kidney, and pancreas.

CC -1- DEVELOPMENTAL STAGE: Detectable at early embryonic ages.

CC -1- DOMAIN: The cytoplasmic domain may be involved in the regulation of trafficking and proteolytic processing. Regulation of the proteolytic processing involves initial intracellular domain dimerization (by similarity).

CC -1- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like domain.

CC -1- PTM: Proteolytic cleavage close to the plasma membrane on the external face leads to the release of the soluble growth factor form.

CC -1- PTM: Extensive glycosylation precedes the proteolytic cleavage (by similarity).

CC -1- DISEASE: Involved in a rare t(8;11) chromosomal translocation that fuses the 5' end of OD24 to NRG1 (isoform 8). The product of this translocation was first thought to be an alternatively spliced isoform, called gamma-hergulin. Gamma-hergulin is a soluble activating ligand for the ERBB2-ERBB3 receptor complex and acts as an autocrine growth factor in a specific breast cancer cell line (MDA-MB-175). Not detected in breast carcinoma samples, including ductal, lobular, medullary, and mucinous histological types, neither in other breast cancer cell lines.

CC -1- SIMILARITY: Contains 1 EGF-like domain.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.

CC -1- SIMILARITY: Belongs to the neuregulin family.

CC

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC

```
-----
CC  EMBL; M94165; AAA58638.1; -
DR  EMBL; M94166; AAA58639.1; -
DR  EMBL; M94167; AAA58640.1; -
DR  EMBL; M94168; AAA58641.1; -
DR  EMBL; U02325; AAA19950.1; -
DR  EMBL; U02326; AAA19951.1; -
DR  EMBL; U02327; AAA19952.1; -
DR  EMBL; U02328; AAA19953.1; -
DR  EMBL; U02329; AAA19954.1; -
DR  EMBL; U02330; AAA19955.1; -
DR  EMBL; L12260; AAB59622.1; -

Query Match      51.4%; Score 18; DB 1; Length 639;
Best Local Similarity 30.8%; Pred. NO. 6.4e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy  4 YXXXXXXXXXXSP 16
Db  423 YVSAMTTPARMSP 435
```

Search completed: October 5, 2004, 16:07:30  
Job time : 13.1781 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:56:41 ; Search time 62.2466 Seconds  
(without alignments)  
81.102 Million cell updates/sec

Title: US-09-973-473A-23

Perfect score: 35

Sequence: 1 XXXYXXXXXXXRXSP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	60.0	595	3 Q9P5R8	Q9P5R8 neurospora
2	21	60.0	1126	5 Q9VGK5	Q9VGK5 drosophila
3	20	57.1	115	17 Q9AU7	Q9AU7 aeropyrum p
4	20	57.1	145	16 Q8DWO	Q8DWO streptococ
5	20	57.1	148	16 Q8E2R2	Q8E2R2 streptococ
6	20	57.1	334	11 Q8R295	Q8R295 mus musculu
7	20	57.1	721	4 Q9V608	Q9V608 homo sapien
8	20	57.1	923	5 Q8IQ88	Q8IQ88 drosophila
9	20	57.1	1243	13 Q9W676	Q9W676 brachydanio
10	19	54.3	177	10 Q7XR2	Q7XR2 oryza sativ
11	19	54.3	187	16 Q7T2H6	Q7T2H6 mycobacteri
12	19	54.3	248	10 Q8L4I2	Q8L4I2 oryza sativ
13	19	54.3	329	12 Q84874	Q84874 human polio
14	19	54.3	337	5 Q9GSD9	Q9GSD9 strongyloce
15	19	54.3	395	12 Q69369	Q69369 cercopithec
16	19	54.3	426	12 Q84873	Q84873 human polio

17	19	54.3	455	3 Q9P7I7	Q9P7I7 schizosacch
18	19	54.3	468	16 O53958	O53958 mycobacteri
19	19	54.3	486	4 Q9HA78	Q9HA78 homo sapien
20	19	54.3	486	4 Q96E33	Q96E33 homo sapien
21	19	54.3	486	4 Q8WZ17	Q8WZ17 homo sapien
22	19	54.3	486	4 Q96RP7	Q96RP7 homo sapien
23	19	54.3	498	4 Q8N3P7	Q8N3P7 homo sapien
24	19	54.3	502	10 Q7XKQ8	Q7XKQ8 oryza sativ
25	19	54.3	567	12 O22814	O22814 arabidopsis
26	19	54.3	575	12 Q84872	Q84872 human polio
27	19	54.3	694	16 Q8VJW0	Q8VJW0 mycobacteri
28	19	54.3	729	3 Q8J1I7	Q8J1I7 cryptococcu
29	19	54.3	729	3 Q8J0Z7	Q8J0Z7 cryptococcu
30	19	54.3	729	3 Q8J0X9	Q8J0X9 cryptococcu
31	19	54.3	730	3 Q8J0V9	Q8J0V9 cryptococcu
32	19	54.3	1170	5 Q95T15	Q95T15 drosophila
33	19	54.3	1990	5 Q9U800	Q9U800 drosophila
34	19	54.3	1995	5 Q9W244	Q9W244 drosophila
35	19	54.3	2209	12 Q8QXN9	Q8QXN9 human polio
36	19	54.3	2209	12 Q912A8	Q912A8 human polio
37	19	54.3	2209	12 Q8B3S2	Q8B3S2 poliovirus
38	19	54.3	2221	12 Q84865	Q84865 human polio
39	18	51.4	81	16 Q8YUL4	Q8YUL4 anabaena sp
40	18	51.4	152	3 Q8TG33	Q8TG33 diaporthes p
41	18	51.4	153	16 Q7UFJ7	Q7UFJ7 rhodopirell
42	18	51.4	167	6 Q862K0	Q862K0 bos taurus
43	18	51.4	175	2 Q8KSG3	Q8KSG3 pseudomonas
44	18	51.4	176	16 Q88H86	Q88H86 pseudomonas
45	18	51.4	180	5 Q86M01	Q86M01 mytilus tro

#### ALIGNMENTS

#### RESULT 1

Q9P5R8 ID Q9P5R8 PRELIMINARY; PRT; 595 AA.  
AC Q9P5R8;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Related to SLY41 protein.  
GN B8820.070.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL355933; CAB91454.2; -  
DR PIR; T49653; T49653.  
DR InterPro; IPR001092; HLH basic.  
DR PROSITE; PS00038; HLH\_1.  
SQ SEQUENCE 595 AA; 64798 MW; 24BCB071E278C4CB CRC64;

Query Match 60.0%; Score 21; DB 3; Length 595;  
Best Local Similarity 30.8%; Pred. No. 3.7e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXXXXSP 16  
Db 4 YTASTGSSLRASP 16

#### RESULT 2

Q9VGK5

ID	Q9VGK5	PRELIMINARY;	PRT;	1126 AA.
AC	Q9VGK5; Q9VGK6;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	CG14713 protein.			
GN	CG14713 OR CG18479.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Berkley;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers J.Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,			
RA	Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hootin D., Houston K.A., Howland T.J., Wei M.-H., Idegwam C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,			
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,			
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,			
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;			
RT	"The genome sequence of Drosophila melanogaster.";			
RL	Science 287:2185-2195(2000).			
DR	EMBL; AE003692; AAF54672.1; -			
DR	FlyBase; FBgn0037928; CG14713.			
KW	Hypothetical protein.			
SQ	SEQUENCE 1126 AA; 119908 MW; 76411146615AE74A CRC64;			
Query Match	60.0%;	Score 21;	DB 5;	Length 1126;
Best Local Similarity	30.8%;	Pred No. 6.8e+02;		
Matches	4;	Conservative 0;	Mismatches 9;	Indels 0; Gaps 0;
Qy	4 YXXXXXXXRXSP 16			
Db	251 YSSTTESSRSRP 263			
RESULT 3				
Q9YAU7	PRELIMINARY;	PRT;	115 AA.	
ID	Q9YAU7			
AC	Q9YAU7;			
DT	01-NOV-1999 (TrEMBLrel. 12, Created)			

```

DR TIGRFAMs; TIGR01529; argR_whole; 1.
KW Complete proteome.
SQ SEQUENCE 145 AA; 16135 MW; AEF6B50CBE8B3797 CRC64;

Query Match 57.1%; Score 20; DB 16; Length 145;
Best Local Similarity 30.8%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 62 YSLSTATSNRFSF 74

RESULT 5
Q8E2R2 PRELIMINARY; PRT; 148 AA.
AC Q8E2R2;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN GBS2055.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEM316 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaser P., Rubinok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Laliou L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766856; CAD47714.1; -.
DR SAGAList; gbs2055; -.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001669; Arg_repress.
DR Pfam; PF01316; Arg_repressor; 1.
DR Pfam; PF02863; Arg_repressor C; 1.
DR PRINTS; PR01467; ARGREPRESSOR.
DR ProDom; PD007402; Arg_repress; 1.
DR TIGRFAMs; TIGR01529; argR whole; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 148 AA; 16431 MW; 7478833755242200 CRC64;

Query Match 57.1%; Score 20; DB 16; Length 148;
Best Local Similarity 30.8%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 65 YSLSTATSNRFSF 77

RESULT 6
Q8R295 PRELIMINARY; PRT; 334 AA.
AC Q8R295;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Vomeronasal receptor VIRF4.
GN VIRF4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

DR TIGRFAMs; TIGR01529; argR_whole; 1.
KW Complete proteome.
SQ SEQUENCE 145 AA; 16135 MW; AEF6B50CBE8B3797 CRC64;

Query Match 57.1%; Score 20; DB 16; Length 145;
Best Local Similarity 30.8%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 62 YSLSTATSNRFSF 74

RESULT 5
Q8E2R2 PRELIMINARY; PRT; 148 AA.
AC Q8E2R2;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN GBS2055.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEM316 / Serotype III;
RX MEDLINE=2242508; PubMed=12354221;
RA Glaser P., Rubinok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Laliou L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766856; CAD47714.1; -.
DR SAGAList; gbs2055; -.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001669; Arg_repress.
DR Pfam; PF01316; Arg_repressor; 1.
DR Pfam; PF02863; Arg_repressor C; 1.
DR PRINTS; PR01467; ARGREPRESSOR.
DR ProDom; PD007402; Arg_repress; 1.
DR TIGRFAMs; TIGR01529; argR whole; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 148 AA; 16431 MW; 7478833755242200 CRC64;

Query Match 57.1%; Score 20; DB 16; Length 148;
Best Local Similarity 30.8%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 65 YSLSTATSNRFSF 77

RESULT 6
Q8R295 PRELIMINARY; PRT; 334 AA.
AC Q8R295;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Vomeronasal receptor VIRF4.
GN VIRF4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=129X1/SVJ;
RX MEDLINE=21676859; PubMed=11802169;
RA Rodriguez I., Del Punta K., Rothman A., Ishii T., Mombaerts P.;
RT "Multiple new and isolated families within the mouse superfamily of
RT V1r vomeronasal receptors.";
RL Nat. Neurosci. 5:134-140 (2002).
DR EMBL; AF085515; AAL47920.1; -.
DR MGD; MGI:2159638; V1rf4.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016503; F:pheromone receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signaling; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR InterPro; IPR004072; Vmron_receptor1.
DR Pfam; PF03402; VIR; 1.
DR PRINTS; PR01534; VOMERONASL1R.
DR PROSITE; PS50262; G_PROTEIN_RECEP_FL_2; 1.
KW Receptor.
SQ SEQUENCE 334 AA; 37794 MW; CF6ABC97EB9F47E7 CRC64;

Query Match 57.1%; Score 20; DB 11; Length 334;
Best Local Similarity 30.8%; Pred. No. 4.4e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 246 YMSRTSLSSRSPP 258

RESULT 7
Q9Y608 PRELIMINARY; PRT; 721 AA.
AC Q9Y608;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE LRR FLI-1 interacting protein 2.
GN LRRFIP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=99296829; PubMed=10366446;
RA Fong K.S.K., de Coust H.G.; with the leucine-rich repeat domain of
RT human flightless-1 identified by the yeast two-hybrid system.";
RL Genomics 58:146-157 (1999).
DR EMBL; AF115509; AAD41257.1; -.
DR PIR; T50611; T50611.
DR Genew; HGNC:6703; LRRFIP2.
DR GO; GO:0030275; F:LRR-domain binding; NAS.
SQ SEQUENCE 721 AA; 82171 MW; 94B18B274656CB0B CRC64;

Query Match 57.1%; Score 20; DB 4; Length 721;
Best Local Similarity 30.8%; Pred. No. 9.3e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 225 YVSSRISSARSSP 237

RESULT 8
Q8IQ88 PRELIMINARY; PRT; 923 AA.
AC Q8IQ88;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

```

```

DE CG14837-PB.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazee R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA April J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacht J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodagat, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J.W., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celniker S.E.,
RA Ciamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003558; AAN2044.1; -.
DR FlyBase; FBN0035797; CG14837.
SQ SEQUENCE 923 AA; 101998 MW; 1380CF09D79CB679 CRC64;
Query Match 57.1%; Score 20; DB 5; Length 923;
Best Local Similarity 30.8%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 4 YXXXXXXXRXSP 16
Db 864 YTASLLSSRTSP 876
RESULT 9
Q9W6T6 PRELIMINARY; PRT; 1243 AA.
AC Q9W6T6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Patched-2 protein.
GN PTC2 OR PTC1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Lewis K.E., Concordet J.P., Ingham P.W.;
RT "Characterisation of the second ptc gene in zebrafish."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ007742; CAB39726.1; -.
DR ZFIN; ZDB-GENE-980526-196; ptc2.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008158; P:hedghog receptor activity; IEA.
DR InterPro; IPR003392; Patched_rec.
DR InterPro; IPR004766; Patched.
DR InterPro; IPR000731; SSD_5TM.
DR Pfam; PF02460; Patched; 1.
DR TIGRFAMs; TIGR00918; 2A060602; 1.
DR PROSITE; PS0156; SSD; 1.
DR SEQUENCE 1243 AA; 138548 MW; FAF459DB0C91371 CRC64;
Query Match 57.1%; Score 20; DB 13; Length 1243;
Best Local Similarity 30.8%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 4 YXXXXXXXRXSP 16
Db 620 YADSSADSSRYSP 632
RESULT 10
Q7XR2 PRELIMINARY; PRT; 177 AA.
AC Q7XR2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNEA000620.10 protein.
DE OSJNEA000620.10.
GN OSJNEA000620.10.

```

```

OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,
RA Huang V.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,
RA Chen L., Fan D.L., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,
RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,
RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
RA Zhang R.Q., Guan J.P., Hong G.P.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL606592; CAEO2745.1; --
SQ SEQUENCE 177 AA; 19928 MW; FE6B1BCB7447D53A CRC64;

Query Match 54.3%; Score 19; DB 10; Length 177;
Best Local Similarity 30.8%; Pred. No. 5e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 64 YFFSHATQTRTSP 76

RESULT 11
ID Q7TZH6 PRELIMINARY; PRT; 187 AA.
AC Q7TZH6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PPE family protein.
GN PPE33A OR MB1838.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsenpe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL: BX248340; CAD94541.1; --
KW Complete proteome.
SQ SEQUENCE 187 AA; 18762 MW; AD02BF52DBFEE29C CRC64;

Query Match 54.3%; Score 19; DB 16; Length 187;
Best Local Similarity 23.1%; Pred. No. 5.3e+02;
Matches 3; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 154 YASSATASRLTP 166

RESULT 12
ID Q8L412 PRELIMINARY; PRT; 248 AA.
AC Q8L412;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P0551A11.8 protein (OJ1116_C07.8 protein).

```

```

GN P0551A11.8 OR OJ1116_C07.8.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0551A11."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
clone:OJ1116_C07."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003934; BAB92816.1; --
DR EMBL: AP004253; BAC00691.1; --
DR Gramene: Q8L412; --
DR GO: GO:0046872; F:metal ion binding; IEA.
DR GO: GO:0030001; P:metal ion transport; IEA.
DR InterPro: IPR006121; HeavyMe_transpt.
DR Pfam: PF00403; HMA; 1.
DR PROSITE: PS50846; HMA; 2; 1.
DR PROSITE: PS50846; HMA; 2; 1.
SQ SEQUENCE 248 AA; 26421 MW; 78A4D2C274333F6B CRC64;

Query Match 54.3%; Score 19; DB 10; Length 248;
Best Local Similarity 30.8%; Pred. No. 7e+02;
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 YXXXXXXXRXSP 16
Db 180 YYAHSYSHRASP 192

RESULT 13
Q84874 PRELIMINARY; PRT; 329 AA.
ID Q84874
AC Q84874;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Human poliovirus 1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=12080;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mahoney strain;
RX MEDLINE=81220953; PubMed=6264310;
RA Kitamura N., Semler B.L., Rothberg P.G., Larsen G.R., Adler C.J.,
RA Dotner A.J., Emini E.A., Hanecak R., Lee J.J., Der Wurf S.,
RA Anderson C.W., Wimmer E.;
RT "Primary structure, gene organization and polypeptide expression of
poliovirus RNA."
RL Nature 291:547-553(1981).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Mahoney strain;
RX MEDLINE=82060159; PubMed=6272282;
RA Racanelli V.R., Baltimore D.;
RT "Molecular cloning of poliovirus cDNA and determination of the
complete nucleotide sequence of the viral genome."
RL Proc. Natl. Acad. Sci. U.S.A. 78:4887-4891(1981).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Mahoney strain;

```

RX MEDLINE=82216986; PubMed=6283138;  
RA Emini E.A., Elzinga M., Wimmer E.;  
RT "Carboxy-Terminal Analysis of Poliovirus Proteins: Termination of  
RT Poliovirus RNA Translation and Location of Unique Poliovirus  
RT Polypeptide Cleavage Sites.";  
RL J. Virol. 42:194-199(1982).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Mahoney strain;  
RX MEDLINE=82242310; PubMed=6284987;  
RA Dörner A.J., Dörner L.F., Larsen G.R., Wimmer E., Anderson C.W.;  
RT "Identification of the Initiation Site of Poliovirus Polyprotein  
RT Synthesis.";  
RL J. Virol. 42:1017-1028(1982).  
DR EMBL; V01148; CAA24455.1; -  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0003724; F:RNA helicase activity; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR InterPro; IPR003593; AAA ATPase.  
DR InterPro; IPR004004; Calici\_pol\_hel.  
DR Efam; PF00910; RNA helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUSNS.  
DR SMART; SM00382; AAA; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 329 AA; 37557 MW; EEECB9320F8A872 CRC64;

Query Match 54.3%; Score 19; DB 12; Length 329;  
Best Local Similarity 30.8%; Pred. No. 9.2e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
DB 217 YVLASTNSRISP 229

RESULT 14  
Q9GSD9 PRELIMINARY; PRT; 337 AA.  
AC Q9GSD9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Transcriptional repressor Krl.  
OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;  
OC Strongylocentrotus.  
OX NCBI\_TaxID=7668;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21067888; PubMed=11152635;  
RA Howard E.W., Newman L.A., Oleksyn D.W., Angerer R.C., Angerer L.M.;  
RT "SpKrl: a direct target of beta-catenin regulation required for  
RT endoderm differentiation in sea urchin embryos.";  
RL Development 128:365-375(2001).  
DR EMBL; AE314167; ARG31160.1; -  
DR HSSP; P25490; IUBD.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf-C2H2; 4.  
DR SMART; SM00355; Znf\_C2H2; 4.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 4.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 337 AA; 37297 MW; B8E713F79DB88F46 CRC64;

Query Match 54.3%; Score 19; DB 5; Length 337;  
Best Local Similarity 30.8%; Pred. No. 9.4e+02;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
DB 217 YVLASTNSRISP 229

Db 147 YNSDTEGSRSP 159  
RESULT 15  
Q9369 PRELIMINARY; PRT; 395 AA.  
AC Q9369;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Glycoprotein gp.  
OS Cercopithecine herpesvirus 2.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae.  
OX NCBI\_TaxID=10317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93298054; PubMed=8390827;  
RA Eberle R., Zhang M., Black D.;  
RT "Gene mapping and sequence analysis of the unique short region of the  
RT simian herpesvirus SA 8 genome.";  
RL Arch. Virol. 130:391-411(1993).  
DR EMBL; AF449714; AAA46178.1; -  
DR InterPro; IPR002896; Herpes\_glycop\_D.  
DR InterPro; IPR007110; Ig-like.  
DR Pfam; PF01537; Herpes\_glycop\_D; 1.  
SQ SEQUENCE 395 AA; 42424 MW; B725333E106417ED CRC64;

Query Match 54.3%; Score 19; DB 12; Length 395;  
Best Local Similarity 30.8%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 YXXXXXXXRXSP 16  
DB 26 YVADRALARASP 38

Search completed: October 5, 2004, 16:12:27  
Job time : 65.2466 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:06 ; Search time 82.1918 Seconds  
(without alignments)  
55.003 Million cell updates/sec

Title: US-09-973-473a-24  
Perfect score: 43  
Sequence: I XXXYXXFLXXRXSP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	28	65.1	378	4	ABG12444	ABG12444	Novel hum
2	28	65.1	727	5	AAM47989	Aam47989	Simian TT
3	27	62.8	437	2	AAY37592	Aay37592	Amino aci
4	26	60.5	16	3	AAB11105	Aab11105	B. mori e
5	26	60.5	16	3	AAB11094	Aab11094	Human eIF
6	26	60.5	16	3	AAB11090	Aab11090	eIF-4E re
7	26	60.5	16	3	AAB11086	Aab11086	eIF-4E re
8	26	60.5	16	4	AAB84410	Aab84410	4E-bindin
9	26	60.5	16	4	AAB84394	Aab84394	4E-bindin
10	26	60.5	16	4	AAB84398	Aab84398	4E-bindin
11	26	60.5	120	2	AAY94275	Aay94275	Human eIF
12	26	60.5	120	3	AAY96148	Aay96148	Human eIF
13	26	60.5	120	5	ABG31602	ABG31602	Human eIF
14	25	58.1	16	3	AAB11093	Aab11093	Human eIF
15	25	58.1	16	3	AAB11104	Aab11104	H. roretz
16	25	58.1	16	3	AAB11099	Aab11099	Chicken e
17	25	58.1	16	3	AAB11085	Aab11085	eIF-4E re
18	25	58.1	16	3	AAB11088	Aab11088	eIF-4E re
19	25	58.1	16	3	AAB11089	Aab11089	eIF-4E re
20	25	58.1	16	4	AAB84404	Aab84404	4E-bindin
21	25	58.1	16	4	AAB84393	Aab84393	4E-bindin
22	25	58.1	16	4	AAB84409	Aab84409	4E-bindin
23	25	58.1	16	4	AAB84396	Aab84396	4E-bindin
24	25	58.1	16	4	AAB84397	Aab84397	4E-bindin
25	25	58.1	117	4	ABB59365	Abb59365	Drosophil

26	25	58.1	117	5	ABB57347	Abb57347	Mouse isc
27	25	58.1	117	5	AAM47601	Aam47601	Drosophil
28	25	58.1	117	7	ADE62052	Ade62052	Rat Prote
29	25	58.1	117	7	ADE62056	Ade62056	Rat Prote
30	25	58.1	118	2	AAW94274	Aaw94274	Human eIF
31	25	58.1	118	3	AAW96147	Aay96147	Human eIF
32	25	58.1	118	5	ABG31601	ABG31601	Human eIF
33	25	58.1	118	5	ABB97146	Abb97146	Human tum
34	25	58.1	118	7	ADD18622	Adi18622	Human dis
35	25	58.1	118	7	ADE62054	Ade62054	Human Pro
36	25	58.1	118	7	ADE62058	Ade62058	Human Pro
37	25	58.1	138	3	AAB43719	Aab43719	Human can
38	25	58.1	964	7	ADC86169	Adc86169	Human GPC
39	24	55.8	16	3	AAB11091	Aab11091	eIF-4E re
40	24	55.8	16	3	AAB11106	Aab11106	S. mansoni
41	24	55.8	16	4	AAB84399	Aab84399	4E-bindin
42	24	55.8	16	4	AAB84411	Aab84411	4E-bindin
43	24	55.8	23	2	AAW67943	Aaw67943	Fragment
44	24	55.8	64	2	AAW67942	Aaw67942	Fragment
45	24	55.8	79	4	AAW24442	Aam24442	Human EST

ALIGNMENTS

RESULT 1  
ABG12444  
ID ABG12444 standard; protein; 378 AA.  
XX  
AC ABG12444;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #12435.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
XX 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS76631.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
XX  
XX Claim 20; SEQ ID NO 42803; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping.  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal  
XX activity of (II) or to treat disease states involving (II). (II) is  
XX useful for generating antibodies against it, detecting or quantitating a  
XX polypeptide in tissue, as molecular weight markers and as a food  
XX supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 378 AA;

Query Match 65.1%; Score 28; DB 4; Length 378;  
 Best Local Similarity 46.2%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXXXSP 16  
 DB 25 YGGDFLRARRSSP 37

RESULT 2  
 AAM47989  
 ID AAM47989 standard; protein; 727 AA.

XX AAM47989;  
 XX 29-AUG-2003 (revised)  
 DT 07-AUG-2003 (revised)  
 DT 07-MAR-2002 (first entry)  
 XX Simian TTV CH71 SEQ ID NO 12.  
 XX Simian TTV; virus; diagnosis; hepatitis; CH71.

XX Viruses.  
 XX WO200185771-A1.

XX 15-NOV-2001.  
 XX 11-MAY-2001; 2001WO-JP003954.  
 XX 11-MAY-2000; 2000JP-00137894.  
 XX 08-NOV-2000; 2000JP-00340614.  
 XX (EISA ) EISAI CO LTD.

XX Abe K;  
 XX WPI; 2002-097552/13.  
 DR N-PSDB; ABA05995.

XX DNA and proteins for diagnosis of non-A, non-B, non-C hepatitis.  
 XX Disclosure; Page 28-29; 37pp; Japanese.

XX The invention relates to an isolated nucleic acid, comprising a 3899  
 CC (S1=ABA05987) or 3322 (S2=ABA05995) nucleotide sequence fully defined in  
 CC the specification useful for diagnosis of non-A, non-B, non-C hepatitis.  
 CC The present sequence is that of a TTV polypeptide, useful to the  
 CC invention. (Updated on 07-AUG-2003 to correct OS field.) (Updated on 29-  
 CC AUG-2003 to standardise OS field)

XX Sequence 727 AA;

Query Match 65.1%; Score 28; DB 5; Length 727;  
 Best Local Similarity 38.5%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXXXSP 16

DB 380 YSIFLSSGRSNP 392

RESULT 3  
 AAY37592

ID AAY37592 standard; protein; 437 AA.

XX AAY37592;

DT 07-OCT-1999 (first entry)

XX Amino acid sequence of a Chlamydia trachomatis protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB001939.

XX 28-NOV-1997; 97FR-00015041.

XX 17-DEC-1997; 97FR-00016034.

XX 04-NOV-1998; 98US-0107077P.

XX (GEST ) GENSET.

XX Griffais R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis.

XX Disclosure; Page 1241; 1755pp; English.

XX AAY36754-V37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perinephritis, bartholinitis; pneumopathy in breast feeding infants; and  
 CC venereal lymphogranulomatosis. The polypeptides of the invention may be  
 CC of use in treating these diseases

XX Sequence 437 AA;

Query Match 62.8%; Score 27; DB 2; Length 437;  
 Best Local Similarity 46.2%; Pred. No. 30;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXXXSP 16

DB 299 YOEFLKKRTSP 311

RESULT 4

AAB11105

ID AAB11105 standard; peptide; 16 AA.

XX AAB11105;

XX 16-FEB-2001 (first entry)

DE B. mori eIF-4E recognition motif peptide 4E-BP.



XX Eukaryotic initiation factor; eIF4E; transgenic animal; antidiabetic;  
 KW knockout mutation; 4E-BP1; glucose metabolism; fat metabolism; anorectic;  
 KW drug; anti-obesity; anti-fat deposition; anti-metabolic.  
 XX Bombyx mori.  
 XX WO200060932-A1.  
 PN 19-OCT-2000.  
 XX 07-APR-2000; 2000WO-CA000388.  
 XX 09-APR-1999; 99US-0128559P.  
 PR 02-FEB-2000; 2000US-0179743P.  
 XX (UYMC-) UNIV MCGILL.  
 XX Sonenberg N, Tremblay M, Tsukiyama-Kohara K;  
 PI WPI; 2000-672657/65.  
 DR Non-human transgenic animal useful as model for studying lipid and  
 XX glucose metabolism, has germ and somatic cells containing knockout  
 PT mutation in DNA encoding eukaryotic initiation factor 4E-binding protein  
 PT 1.  
 XX Disclosure; Fig 7; 80pp; English.  
 XX This invention describes a novel non-human transgenic animal (I) whose  
 CC germ cells and somatic cells contain a knockout mutation in DNA encoding  
 CC a 4E-BP1 (a member of eukaryotic initiation factor 4E (eIF-4E)-binding  
 CC protein 1) shows a phenotype of an altered glucose and/or fat metabolism  
 CC when compared to a control animal. The products of the invention have  
 CC antidiabetic and anorectic activity. The non-human transgenic animals are  
 CC useful as models for the investigation of lipid and glucose metabolism,  
 CC energy homeostasis and associated diseases. The transgenic animals are  
 CC also useful for selection and identification of modulators of the  
 CC expression and/or activity of 4E-BP1 and for screening drugs to isolate  
 CC therapeutic agents which are useful as anti-obesity, anti-fat deposition  
 CC disorder, anti-diabetes and anti-metabolic diseases associated with fat  
 CC deposition disorders. The effect of the disruption of 4E-BP1 in the  
 CC knockout mice demonstrates that an alteration of 4E-BP1 can modulate fat  
 CC tissue growth, glucose metabolism and weight gain in an animal  
 XX Sequence 16 AA;  
 SQ  
 Query Match 60.5%; Score 26; DB 3; Length 16;  
 Best Local Similarity 38.5%; Pred. No. 2;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
 QY 4 YXXXFLXXXRXSP 16  
 Db | | | | | | | |  
 4 YERSFMLSRLQSP 16  
 RESULT 5  
 AAB11094  
 ID AAB11094 standard; peptide; 16 AA.  
 AC AAB11094;  
 XX 16-FEB-2001 (first entry)  
 DT Human eIF-4E recognition motif peptide h4E-BP2.  
 DE Eukaryotic initiation factor; eIF4E; transgenic animal; antidiabetic;  
 KW knockout mutation; 4E-BP1; glucose metabolism; fat metabolism; anorectic;  
 KW drug; anti-obesity; anti-fat deposition; anti-metabolic.  
 XX Homo sapiens.  
 OS WO200060932-A1.  
 PN

XX 19-OCT-2000.  
 XX 07-APR-2000; 2000WO-CA000388.  
 XX 09-APR-1999; 99US-0128559P.  
 PR 02-FEB-2000; 2000US-0179743P.  
 XX (UYMC-) UNIV MCGILL.  
 XX Sonenberg N, Tremblay M, Tsukiyama-Kohara K;  
 PI WPI; 2000-672657/65.  
 DR Non-human transgenic animal useful as model for studying lipid and  
 XX glucose metabolism, has germ and somatic cells containing knockout  
 PT mutation in DNA encoding eukaryotic initiation factor 4E-binding protein  
 PT 1.  
 XX Disclosure; Fig 7; 80pp; English.  
 XX This invention describes a novel non-human transgenic animal (I) whose  
 CC germ cells and somatic cells contain a knockout mutation in DNA encoding  
 CC a 4E-BP1 (a member of eukaryotic initiation factor 4E (eIF-4E)-binding  
 CC protein 1) shows a phenotype of an altered glucose and/or fat metabolism  
 CC when compared to a control animal. The products of the invention have  
 CC antidiabetic and anorectic activity. The non-human transgenic animals are  
 CC useful as models for the investigation of lipid and glucose metabolism,  
 CC energy homeostasis and associated diseases. The transgenic animals are  
 CC also useful for selection and identification of modulators of the  
 CC expression and/or activity of 4E-BP1 and for screening drugs to isolate  
 CC therapeutic agents which are useful as anti-obesity, anti-fat deposition  
 CC disorder, anti-diabetes and anti-metabolic diseases associated with fat  
 CC deposition disorders. The effect of the disruption of 4E-BP1 in the  
 CC knockout mice demonstrates that an alteration of 4E-BP1 can modulate fat  
 CC tissue growth, glucose metabolism and weight gain in an animal  
 XX Sequence 16 AA;  
 SQ  
 Query Match 60.5%; Score 26; DB 3; Length 16;  
 Best Local Similarity 46.2%; Pred. No. 2;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 4 YXXXFLXXXRXSP 16  
 Db | | | | | | | |  
 4 YDRKFLDDRNSP 16  
 RESULT 6  
 AAB11090  
 ID AAB11090 standard; peptide; 16 AA.  
 AC AAB11090;  
 XX 16-FEB-2001 (first entry)  
 DT eIF-4E recognition motif peptide bm4E-BP.  
 DE Eukaryotic initiation factor; eIF4E; transgenic animal; antidiabetic;  
 KW knockout mutation; 4E-BP1; glucose metabolism; fat metabolism; anorectic;  
 KW drug; anti-obesity; anti-fat deposition; anti-metabolic.  
 XX Unidentified.  
 OS WO200060932-A1.  
 PN 19-OCT-2000.  
 XX 07-APR-2000; 2000WO-CA000388.  
 XX 09-APR-1999; 99US-0128559P.  
 PR 02-FEB-2000; 2000US-0179743P.  
 XX

PA (UYMC-) UNIV MCGILL.  
XX Sonenberg N, Tremblay M, Tsukiyama-Kohara K;  
XX WPI; 2000-672657/65.  
XX  
XX Non-human transgenic animal useful as model for studying lipid and  
PT glucose metabolism, has germ and somatic cells containing knockout  
PT mutation in DNA encoding eukaryotic initiation factor 4E-binding protein  
PT 1.  
XX  
XX Disclosure; Fig 6; 80pp; English.  
XX  
XX This invention describes a novel non-human transgenic animal (I) whose  
CC germ cells and somatic cells contain a knockout mutation in DNA encoding  
CC a 4E-BP1 (a member of eukaryotic initiation factor 4E (eIF-4E)-binding  
CC protein 1) shows a phenotype of an altered glucose and/or fat metabolism  
CC when compared to a control animal. The products of the invention have  
CC antidiabetic and anorectic activity. The non-human transgenic animals are  
CC useful as models for the investigation of lipid and glucose metabolism,  
CC energy homeostasis and associated diseases. The transgenic animals are  
CC also useful for selection and identification of modulators of the  
CC expression and/or activity of 4E-BP1 and for screening drugs to isolate  
CC therapeutic agents which are useful as anti-obesity, anti-fat deposition  
CC disorder, anti-diabetes and anti-metabolic diseases associated with fat  
CC deposition disorders. The effect of the disruption of 4E-BP1 in the  
CC knockout mice demonstrates that an alteration of 4E-BP1 can modulate fat  
CC tissue growth, glucose metabolism and weight gain in an animal  
XX  
XX Sequence 16 AA;  
XX  
XX Query Match 60.5%; Score 26; DB 3; Length 16;  
XX Best Local Similarity 38.5%; Pred. No. 2;  
XX Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
QY 4 YXXXFLXXRXSP 16  
DB | | | | |  
4 YERSFMSLRQSP 16  
RESULT 7  
AAB11086  
ID AAB11086 standard; peptide; 16 AA.  
XX  
XX AAB11086;  
XX  
XX 16-FEB-2001 (first entry)  
XX  
XX eIF-4E recognition motif peptide h4E-BP2.  
XX  
XX Eukaryotic initiation factor; eIF4E; transgenic animal; antidiabetic;  
XX knockout mutation; 4E-BP1; glucose metabolism; fat metabolism; anorectic;  
XX drug; anti-obesity; anti-fat deposition; anti-metabolic.  
XX  
XX Unidentified.  
XX  
XX WO200060932-A1.  
XX  
XX 19-OCT-2000.  
XX  
XX 07-APR-2000; 2000WO-CA000388.  
XX  
XX 09-APR-1999; 99US-0128559P.  
XX  
XX 02-FEB-2000; 2000US-0179743P.  
XX  
XX (UYMC-) UNIV MCGILL.  
XX  
XX Sonenberg N, Tremblay M, Tsukiyama-Kohara K;  
XX WPI; 2000-672657/65.  
XX  
XX Non-human transgenic animal useful as model for studying lipid and  
PT glucose metabolism, has germ and somatic cells containing knockout  
PT mutation in DNA encoding eukaryotic initiation factor 4E-binding protein  
PT 1.  
XX  
XX Disclosure; Fig 6; 80pp; English.  
XX  
XX This invention describes a novel non-human transgenic animal (I) whose  
CC germ cells and somatic cells contain a knockout mutation in DNA encoding  
CC a 4E-BP1 (a member of eukaryotic initiation factor 4E (eIF-4E)-binding  
CC protein 1) shows a phenotype of an altered glucose and/or fat metabolism  
CC when compared to a control animal. The products of the invention have  
CC antidiabetic and anorectic activity. The non-human transgenic animals are  
CC useful as models for the investigation of lipid and glucose metabolism,  
CC energy homeostasis and associated diseases. The transgenic animals are  
CC also useful for selection and identification of modulators of the  
CC expression and/or activity of 4E-BP1 and for screening drugs to isolate  
CC therapeutic agents which are useful as anti-obesity, anti-fat deposition  
CC disorder, anti-diabetes and anti-metabolic diseases associated with fat  
CC deposition disorders. The effect of the disruption of 4E-BP1 in the  
CC knockout mice demonstrates that an alteration of 4E-BP1 can modulate fat  
CC tissue growth, glucose metabolism and weight gain in an animal  
XX  
XX Sequence 16 AA;  
XX  
XX Query Match 60.5%; Score 26; DB 3; Length 16;  
XX Best Local Similarity 46.2%; Pred. No. 2;  
XX Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 4 YXXXFLXXRXSP 16  
DB | | | | |  
4 YDRKFLDRLNSP 16  
RESULT 8  
AAB84410  
ID AAB84410 standard; peptide; 16 AA.  
XX  
XX AAB84410;  
XX  
XX 22-AUG-2001 (first entry)  
XX  
XX 4E-binding site of a 4E-binding protein.  
XX  
XX 4E-binding site; 4E-binding protein; 4E-BP; 4E-BP1; chemoresistance;  
XX cap-dependent translation initiation repressor; apoptosis;  
XX translation initiation factor; eIF4E; Ras; cancer.  
XX  
XX Bombyx mori.  
XX  
XX WO200140293-A2.  
XX  
XX 07-JUN-2001.  
XX  
XX 01-DEC-2000; 2000WO-CA001465.  
XX  
XX 02-DEC-1999; 99US-0168398P.  
XX  
XX (UYMC-) UNIV MCGILL.  
XX (MINU ) UNIV MINNESOTA.  
XX  
XX Sonenberg N, Gingras A, Polunavsky VA, Bitterman PB;  
XX WPI; 2001-381379/40.  
XX  
XX Inducing apoptosis in a cell in which the apoptosis pathway is inhibited,  
PT useful for treating cancer or highly proliferative cells, comprises  
PT decreasing the amount of eIF4F pre-initiation complex, relieving an  
PT apoptosis block.  
XX  
XX Disclosure; Fig 6; 80pp; English.  
XX  
XX AAB84401-15 represent 4E-binding sites. 4E-binding protein 1 (4E-BP1) is  
CC a repressor of cap-dependent translation initiation, and selectively  
CC activates apoptosis in Ras-transformed fibroblasts and eliminates Ras-

CC induced chemoresistance. The effects of 4E-BP1 are strictly dependent on  
CC its ability to sequester the translation initiation factor eIF4E, thereby  
CC preventing its assembly into an active pre-initiation complex. The  
CC specification describes a method for inducing apoptosis in a cell in  
CC which the apoptosis pathway is inhibited. The method comprises decreasing  
CC the amount of eIF4F pre-initiation complex by sequestration of eIF4E,  
CC thus relieving an apoptosis block. The method is useful for modulating  
CC pro-apoptotic and anti-apoptotic pathways in cells, especially in Ras-  
CC transformed cells. Thus, the method is particularly useful for treating  
CC cancer. The eIF4E sequestering agent, 4E-BP1 or its eIF4E binding portion  
CC is useful for treating high proliferative cells  
XX  
SQ Sequence 16 AA;

Query Match 60.5%; Score 26; DB 4; Length 16;  
Best Local Similarity 38.5%; Pred. No. 2;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXRXSP 16  
| | | | |  
DB 4 YERSFMLSROSP 16

## RESULT 9

AAB84394  
ID AAB84394 standard; peptide; 16 AA.

XX AAB84394;

DT 22-AUG-2001 (first entry)

DE 4E-binding site of a human 4E-binding protein 2.

XX 4E-binding site; 4E-binding protein; 4E-BP; 4E-BP1; chemoresistance;  
KW cap-dependent translation initiation repressor; apoptosis;  
KW translation initiation factor; eIF4E; Ras; Cancer.

OS Homo sapiens.

PN WO200140293-A2.

XX 07-JUN-2001.

PF 01-DEC-2000; 2000WO-CA001465.

PR 02-DEC-1999; 99US-0168398P.

XX (UYMC-) UNIV MCGILL.

PA (MINU ) UNIV MINNESOTA.

XX Sonenberg N, Gingras A, Polunavsky VA, Bitterman PB;

PI WPI; 2001-381379/40.

DR Inducing apoptosis in a cell in which the apoptosis pathway is inhibited,  
XX useful for treating cancer or highly proliferative cells, comprises  
PT decreasing the amount of eIF4F pre-initiation complex, relieving an  
PT apoptosis block.

PS Disclosure; Fig 5; 80pp; English.

XX AAB84393-AAB84400 represent 4E-binding sites from 4E-binding proteins (4E  
CC -BPs). 4E-BP1 is a repressor of cap-dependent translation initiation, and  
CC selectively activates apoptosis in Ras-transformed fibroblasts and  
CC eliminates Ras-induced chemoresistance. The effects of 4E-BP1 are  
CC strictly dependent on its ability to sequester the translation initiation  
CC factor eIF4E, thereby preventing its assembly into an active pre-  
CC initiation complex. The specification describes a method for inducing  
CC apoptosis in a cell in which the apoptosis pathway is inhibited. The  
CC method comprises decreasing the amount of eIF4F pre-initiation complex by  
CC sequestration of eIF4E, thus relieving an apoptosis block. The method is  
CC useful for modulating pro-apoptotic and anti-apoptotic pathways in cells,  
CC especially in Ras-transformed cells. Thus, the method is particularly

CC useful for treating cancer. The eIF4E sequestering agent, 4E-BP1 or its  
CC eIF4E binding portion is useful for treating high proliferative cells  
XX  
SQ Sequence 16 AA;

Query Match 60.5%; Score 26; DB 4; Length 16;  
Best Local Similarity 46.2%; Pred. No. 2;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXRXSP 16  
| | | | |  
DB 4 YDRKFLDRLRRNSP 16

## RESULT 10

AAB84398  
ID AAB84398 standard; peptide; 16 AA.

XX AAB84398;

DT 22-AUG-2001 (first entry)

DE 4E-binding site of a 4E-binding protein.

XX 4E-binding site; 4E-binding protein; 4E-BP; 4E-BP1; chemoresistance;  
KW cap-dependent translation initiation repressor; apoptosis;  
KW translation initiation factor; eIF4E; Ras; Cancer.

OS Bombyx mori.

PN WO200140293-A2.

XX 07-JUN-2001.

PF 01-DEC-2000; 2000WO-CA001465.

PR 02-DEC-1999; 99US-0168398P.

XX (UYMC-) UNIV MCGILL.

PA (MINU ) UNIV MINNESOTA.

XX Sonenberg N, Gingras A, Polunavsky VA, Bitterman PB;

PI WPI; 2001-381379/40.

DR Inducing apoptosis in a cell in which the apoptosis pathway is inhibited,  
XX useful for treating cancer or highly proliferative cells, comprises  
PT decreasing the amount of eIF4F pre-initiation complex, relieving an  
PT apoptosis block.

PS Disclosure; Fig 5; 80pp; English.

XX AAB84393-AAB84400 represent 4E-binding sites from 4E-binding proteins (4E  
CC -BPs). 4E-BP1 is a repressor of cap-dependent translation initiation, and  
CC selectively activates apoptosis in Ras-transformed fibroblasts and  
CC eliminates Ras-induced chemoresistance. The effects of 4E-BP1 are  
CC strictly dependent on its ability to sequester the translation initiation  
CC factor eIF4E, thereby preventing its assembly into an active pre-  
CC initiation complex. The specification describes a method for inducing  
CC apoptosis in a cell in which the apoptosis pathway is inhibited. The  
CC method comprises decreasing the amount of eIF4F pre-initiation complex by  
CC sequestration of eIF4E, thus relieving an apoptosis block. The method is  
CC useful for modulating pro-apoptotic and anti-apoptotic pathways in cells,  
CC especially in Ras-transformed cells. Thus, the method is particularly  
CC useful for treating cancer. The eIF4E sequestering agent, 4E-BP1 or its  
CC eIF4E binding portion is useful for treating high proliferative cells  
XX

SQ Sequence 16 AA;

Query Match 60.5%; Score 26; DB 4; Length 16;  
Best Local Similarity 38.5%; Pred. No. 2;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXXXSP 16  
DB 4 YERSFMSLRQSP 16

RESULT 11  
ID AAW94275  
AAW94275 standard; protein; 120 AA.

XX AC AAW94275;  
XX XX  
DT 22-APR-1999 (first entry)  
XX XX  
DE Human eIF-4E-binding protein 4E-BP2.  
XX XX  
KW Translation factor; cellular factor; hormone; inhibitor; screening;  
KW diabetes; eIF-4E; 4E-BP2; binding protein.  
XX XX

OS Homo sapiens.  
PN US5874231-A.  
XX XX

PD 23-FEB-1999.  
XX XX

XX 22-AUG-1994; 94US-00294143.  
XX XX

XX 22-AUG-1994; 94US-00294143.  
XX XX

XX (RIBO-) RIBOGENE INC.  
PA (UYMC-) UNIV MCGILL.  
XX XX

PI Miles VJ, Harford JB, Sonenberg N, Pause A;  
XX XX

DR WPI; 1999-179971/15.  
DR N-PSDB; AAW5654.  
XX XX

XX Screening assay for modulators of hormone activity - based on hormone-  
PT dependent release of translation factor.  
XX XX

PS Example 1; Fig 2; 29pp; English.  
XX XX

CC The invention relates to screening methods for identifying agents that  
CC mimic the activity of a hormone. The method comprises a test agent with  
CC an in vitro system containing a complex comprising a translation factor  
CC sequestered by a cellular component, where the complex responds to  
CC hormone treatment by releasing the translation factor, and determining if  
CC the test agent causes the release of the translation factor by an assay  
CC that directly measures the translation factor and/or the cellular  
CC component or an assay that measures the activity of the released  
CC translation factor by measuring the expression of a reporter polypeptide  
CC from a reporter construct. Inhibitors of a hormone can also be identified  
CC using a similar method. The methods are useful for screening for drugs  
CC useful in the treatment of hormonal disorders, especially diabetes. The  
CC translation factor used is eIF-4E. The present sequence represents a  
CC human eIF-4E-binding protein 4E-BP2 that can be used as a cellular factor  
CC in the method  
XX XX

SQ Sequence 120 AA;  
Query Match 60.5%; Score 26; DB 2; Length 120;  
Best Local Similarity 46.2%; Pred. No. 16;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXXXSP 16  
DB 54 YDRKFLDRNSP 66

RESULT 12  
ID AAY96148  
AAY96148 standard; protein; 120 AA.  
XX XX  
AC AAY96148;

XX 19-DEC-2000 (first entry)  
XX XX  
DE Human eIF-4E binding protein 4E-BP2.  
XX XX

KW Human; 4E-BP2; eukaryotic initiation factor 4E; eIF-4E; binding protein;  
KW translation factor; hormone disorder; therapy; diabetic; antidiabetic;  
KW hypoglycemic.  
XX XX

OS Homo sapiens.  
PN US6111077-A.  
XX XX

PD 29-AUG-2000.  
XX XX

PF 23-FEB-1999; 99US-00256331.  
XX XX

PR 22-AUG-1994; 94US-00294143.  
XX XX

PA (RIBO-) RIBOGENE INC.  
PA (UYMC-) UNIV MCGILL.  
XX XX

PI Harford JB, Miles VJ, Sonenberg N, Pause A;  
XX XX

DR WPI; 2000-571431/53.  
DR N-PSDB; AAA50622.  
XX XX

XX Human cellular component that binds to translation factor eIF-4E causing  
PT modulation of translation in cell, useful for screening agents for  
PT prophylaxis and treatment of hormone disorders such as diabetes.  
XX XX

PS Example 1; Fig 2; 27pp; English.  
XX XX

CC The present sequence is that of human eukaryotic initiation factor 4E  
CC (eIF-4E) binding protein 4E-BP2, as predicted from a cDNA clone (see  
CC AAA50622) isolated from a placenta cDNA library. 4E-BP2 protein binds to  
CC eIF-4E and causes a modulation of translation in a cell in response to a  
CC hormone. Phosphorylation of the protein, e.g. in response to insulin,  
CC causes its release from eIF-4E and a consequent stimulation of  
CC translation. 4E-BP2 shows 56% amino acid identity to 4E-BP1 (see  
CC AAY96147). It can be used in a method for screening for non-hormone  
CC agents potentially useful for treating a hormone disorder. The method  
CC involves contacting a potential agent with a system (e.g. in vitro  
CC translation system) containing a cellular component and a translation  
CC factor. The component and factor interact in an intact normal cell in a  
CC manner responsive to the hormone to cause a modulation of translation in  
CC the cell. The method involves determining whether the agent causes a  
CC modulation of translation by the component and the factor autologous to  
CC that which occurs in intact cells in response to the hormone. The agent  
CC is useful for prophylaxis and treatment of hormone disorders such as  
CC diabetes. The agent is useful as a hormone substitute and also in test  
CC systems to allow an understanding of the action of the hormone. The  
CC improved drugs for treating diabetes mellitus are more convenient to  
CC administer than insulin, avoid the side effects of oral hypoglycemic  
CC drugs and carry a reduced risk of inducing severe hypoglycemia  
XX XX

SQ Sequence 120 AA;

Query Match 60.5%; Score 26; DB 3; Length 120;  
Best Local Similarity 46.2%; Pred. No. 16;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXXXSP 16  
DB 54 YDRKFLDRNSP 66

RESULT 13  
ID ABG31602  
ABG31602 standard; protein; 120 AA.  
XX XX  
AC ABG31602;  
XX XX

DT 05-NOV-2002 (first entry)  
 DE Human eIF-4E-binding protein 4E-BP2.  
 DE Human: eIF-4E-binding protein; eIF-4E; 4E-BP1; 4E-BP2; hormone response;  
 KW hormonal disorder; diabetes; Addison's disease; hyperfunction;  
 KW Cushing's syndrome; pheochromocytoma; multiple endocrine neoplasia;  
 KW polyglandular deficiency syndrome; reproductive function disorder;  
 KW cytostatic; gene therapy; modulation of translation.  
 XX Homo sapiens.  
 XX US6410715-B1.  
 PN 25-JUN-2002.  
 XX 14-JUN-2000; 2000US-00593483.  
 XX 22-AUG-1994; 94US-00294143.  
 PR 23-FEB-1999; 99US-00256331.  
 XX (QUES-) QUESTCOR PHARM INC.  
 PA (UYMC-) UNIV MCGILL.  
 PI Sonenberg N, Pause A, Harford JB, Miles VJ;  
 XX WPI; 2002-582026/62.  
 DR N-PSDB; ASK90700.  
 XX Human nucleic acids encoding 4E-BP 1 and 4E-BP2 proteins which interact  
 PT with eIF-4E translation factor and modulate translation, useful for  
 PT treating hormonal disorders e.g. diabetes and Addison's disease.  
 XX Claim 2; Fig 2; 27pp; English.  
 CC The present invention relates to a new human nucleic acid sequence  
 CC encoding a cellular component that binds to eIF-4E (not defined) and  
 CC causes a modulation of translation in a cell in response to a hormone.  
 CC The invention comprises a coding sequence for the proteins 4E-BP1 or 4E-  
 CC BP2. The nucleic acid encodes a protein that binds to eIF-4E and causes a  
 CC modulation of translation in a cell in response to a hormone. The protein  
 CC encoded by the nucleic acid may be used in controlling gene expression.  
 CC It may be used in this way to treat hormonal disorders (e.g. diabetes,  
 CC Addison's disease) or hyperfunctions (e.g. Cushing's syndrome),  
 CC pheochromocytoma, multiple endocrine neoplasias, polyglandular deficiency  
 CC syndromes, and disorders of reproductive function. The present amino acid  
 CC sequence represents the human eIF-4E-binding protein 4E-BP2 of the  
 CC invention  
 XX  
 SQ Sequence 120 AA;  
 Query Match 60.5%; Score 26; DB 5; Length 120;  
 Best Local Similarity 46.2%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 4 YXXYFLXXXRXSP 16  
 DB 54 YDRKFLDRNSP 66  
 RESULT 14  
 AAB11093  
 ID AAB11093 standard; peptide; 16 AA.  
 AC AAB11093;  
 XX 16-FEB-2001 (first entry)  
 DT Human eIF-4E recognition motif peptide h4E-BP1.  
 DE  
 DE Eukaryotic initiation factor; eIF4E; transgenic animal; antidiabetic;  
 KW knockout mutation; 4E-BP1; glucose metabolism; fat metabolism; anorectic;  
 KW drug; anti-obesity; anti-fat deposition; anti-metabolic.

XX Homo sapiens.  
 OS WO200060932-A1.  
 PN 19-OCT-2000.  
 PD 07-APR-2000; 2000WO-CA000388.  
 PF 09-APR-1999; 99US-0128559P.  
 PR 02-FEB-2000; 2000US-0179743P.  
 XX (UYMC-) UNIV MCGILL.  
 PA Sonenberg N, Tremblay M, Tsukiyama-Kohara K;  
 PI WPI; 2000-672657/65.  
 DR Non-human transgenic animal useful as model for studying lipid and  
 XX glucose metabolism, has germ and somatic cells containing knockout  
 PT mutation in DNA encoding eukaryotic initiation factor 4E-binding protein  
 PT 1.  
 PS Disclosure; Fig 7; 80pp; English.  
 CC This invention describes a novel non-human transgenic animal (I) whose  
 CC germ cells and somatic cells contain a knockout mutation in DNA encoding  
 CC a 4E-BP1 (a member of eukaryotic initiation factor 4E (eIF-4E)-binding  
 CC protein 1) shows a phenotype of an altered glucose and/or fat metabolism  
 CC when compared to a control animal. The products of the invention have  
 CC antidiabetic and anorectic activity. The non-human transgenic animals are  
 CC useful as models for the investigation of lipid and glucose metabolism,  
 CC energy homeostasis and associated diseases. The transgenic animals are  
 CC also useful for selection and identification of modulators of the  
 CC expression and/or activity of 4E-BP1 and for screening drugs to isolate  
 CC therapeutic agents which are useful as anti-obesity, anti-fat deposition  
 CC disorder, anti-diabetes and anti-metabolic diseases associated with fat  
 CC deposition disorders. The effect of the disruption of 4E-BP1 in the  
 CC knockout mice demonstrates that an alteration of 4E-BP1 can modulate fat  
 CC tissue growth, glucose metabolism and weight gain in an animal  
 XX Sequence 16 AA;  
 SQ  
 Query Match 58.1%; Score 25; DB 3; Length 16;  
 Best Local Similarity 46.2%; Pred. No. 4;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 4 YXXYFLXXXRXSP 16  
 DB 4 YDRKFLMECRNSP 16  
 RESULT 15  
 AAB11104  
 ID AAB11104 standard; peptide; 16 AA.  
 AC AAB11104;  
 XX 16-FEB-2001 (first entry)  
 DT H. roretzi eIF-4E recognition motif peptide 4E-BP.  
 DE  
 DE Eukaryotic initiation factor; eIF4E; transgenic animal; antidiabetic;  
 KW knockout mutation; 4E-BP1; glucose metabolism; fat metabolism; anorectic;  
 KW drug; anti-obesity; anti-fat deposition; anti-metabolic.  
 XX Halocynthia roretzi.  
 OS WO200060932-A1.  
 PN 19-OCT-2000.  
 PD 07-APR-2000; 2000WO-CA000388.  
 PF

XX 09-APR-1999; 99US-0128559P.  
PR 02-FEB-2000; 200US-0179743P.  
XX (UYMC-) UNIV MCGILL.  
XX Sonenberg N, Tremblay M, Tsukiyama-Kohara K;  
PI WPI; 2000-672657/65.  
XX  
XX Non-human transgenic animal useful as model for studying lipid and  
PT Glucose metabolism, has germ and somatic cells containing knockout  
PT mutation in DNA encoding eukaryotic initiation factor 4E-binding protein  
PT 1.  
XX  
XX  
PS Disclosure; Fig 7; 80pp; English.  
XX  
XX This invention describes a novel non-human transgenic animal (I) whose  
CC germ cells and somatic cells contain a knockout mutation in DNA encoding  
CC a 4E-BP1 (a member of eukaryotic initiation factor 4E (eIF-4E)-binding  
CC protein 1) shows a phenotype of an altered glucose and/or fat metabolism  
CC when compared to a control animal. The products of the invention have  
CC antidiabetic and anorectic activity. The non-human transgenic animals are  
CC useful as models for the investigation of lipid and glucose metabolism,  
CC energy homeostasis and associated diseases. The transgenic animals are  
CC also useful for selection and identification of modulators of the  
CC expression and/or activity of 4E-BP1 and for screening drugs to isolate  
CC therapeutic agents which are useful as anti-obesity, anti-fat deposition  
CC disorder, anti-diabetes and anti-metabolic diseases associated with fat  
CC deposition disorders. The effect of the disruption of 4E-BP1 in the  
CC knockout mice demonstrates that an alteration of 4E-BP1 can modulate fat  
CC tissue growth, glucose metabolism and weight gain in an animal  
XX  
SQ Sequence 16 AA;

Query Match 58.1%; Score 25; DB 3; Length 16;  
Best Local Similarity 46.2%; Pred. No. 4;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 4 YXXXFLXXXRXSP 16  
Db 4 YDRFLKCRDSP 16

Search completed: October 5, 2004, 16:06:39  
Job time : 85.1918 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:00:17 ; Search time 21.6986 Seconds  
(without alignments)  
38.068 Million cell updates/sec

Title: US-09-973-473a-24

Perfect score: 43

Sequence: 1 XXXXXFLXXXRSP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/pCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	67.4	201	4	US-09-252-991A-21355
2	26	60.5	120	2	US-08-869-733-3
3	26	60.5	120	2	US-08-869-733-4
4	26	60.5	120	4	US-09-215-063-3
5	26	60.5	120	4	US-09-215-063-4
6	26	60.5	120	4	US-10-096-703-3
7	26	60.5	120	4	US-10-096-703-4
8	24	55.8	315	4	US-09-107-532A-5917
9	24	55.8	345	4	US-09-134-000C-5047
10	24	55.8	1447	3	US-09-041-886-25
11	24	55.8	1447	5	PCT-US94-05277-2
12	23	53.5	491	3	US-08-923-454A-14
13	23	53.5	492	1	US-08-350-741-3
14	23	53.5	492	2	US-08-463-875A-3
15	23	53.5	506	4	US-09-489-039A-8921
16	23	53.5	523	3	US-08-651-999A-2
17	23	53.5	523	3	US-09-385-752-2
18	23	53.5	525	4	US-09-253-991A-28266
19	23	53.5	560	2	US-08-948-569A-10
20	23	53.5	560	2	US-09-188-469-10
21	23	53.5	560	3	US-09-397-238A-10
22	23	53.5	607	4	US-09-907-794A-190
23	23	53.5	607	4	US-09-905-125A-190
24	23	53.5	607	4	US-09-902-775A-190
25	23	53.5	630	4	US-09-657-252-4
26	23	53.5	1614	4	US-09-052-469-2
27	23	53.5	1614	4	US-08-422-582-2

28	23	53.5	1614	4	US-09-052-262-2	Sequence 2, Appli
29	23	53.5	4302	3	US-08-658-136-5	Sequence 5, Appli
30	23	53.5	4302	4	US-09-052-469-8	Sequence 8, Appli
31	23	53.5	4302	4	US-08-422-582-8	Sequence 8, Appli
32	23	53.5	4302	4	US-09-052-262-8	Sequence 8, Appli
33	23	53.5	4303	2	US-08-460-751-2	Sequence 2, Appli
34	23	53.5	4339	4	US-09-052-469-6	Sequence 6, Appli
35	23	53.5	4339	4	US-08-422-582-6	Sequence 6, Appli
36	23	53.5	4339	4	US-09-052-262-6	Sequence 6, Appli
37	22	51.2	52	4	US-09-489-847-184	Sequence 184, App
38	22	51.2	74	3	US-09-100-804-33	Sequence 33, Appl
39	22	51.2	80	4	US-09-489-039A-12870	Sequence 12870, A
40	22	51.2	100	2	US-08-869-733-1	Sequence 1, Appli
41	22	51.2	100	4	US-09-215-063-1	Sequence 1, Appli
42	22	51.2	100	4	US-09-673-395A-172	Sequence 172, App
43	22	51.2	100	4	US-10-096-703-1	Sequence 1, Appli
44	22	51.2	263	4	US-09-252-991A-23847	Sequence 23847, A
45	22	51.2	284	4	US-09-252-991A-30334	Sequence 30334, A

## ALIGNMENTS

RESULT 1  
US-09-252-991A-21355  
; Sequence 21355, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21355  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21355

Query Match 67.4%; Score 29; DB 4; Length 201;  
Best Local Similarity 46.2%; Pred. No. 0.91;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy	4	YXXFLXXXRSP	16
Db	34	YAARFLVFSRESP	46

RESULT 2  
US-08-869-733-3  
; Sequence 3, Application US/08869733  
; Patent No. 5955278  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Hawkins, Phillip R.  
; TITLE OF INVENTION: NEW TRANSLATIONAL REGULATOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible







```
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J.
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0310 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-855-0555
/ TELEFAX: 415-845-4166
/ TELEX: <Unknown>
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 120 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 1658516
/ SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-096-703-4
Query Match 60.5%; Score 26; DB 4; Length 120;
Best Local Similarity 46.2%; Pred. No. 4.1;
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXRXSP 16
| | | | |
Db 54 YDRKFLDRNSP 66

RESULT 8
US-09-107-532A-5917
; Sequence 5917, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5917:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ATTORNEY/AGENT INFORMATION:
```

```
/ ORGANISM: Enterococcus faecium
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (8) LOCATION 1...315
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5917:
US-09-107-532A-5917
Query Match 55.8%; Score 24; DB 4; Length 315;
Best Local Similarity 55.6%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXRXSP 16
| | | | |
Db 56 FLASLRESP 64

RESULT 9
US-09-134-000C-5047
; Sequence 5047, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5047
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-134-000C-5047
Query Match 55.8%; Score 24; DB 4; Length 345;
Best Local Similarity 55.6%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXRXSP 16
| | | | |
Db 86 FLASLRESP 94

RESULT 10
US-09-041-886-25
; Sequence 25, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
```

NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-1J 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1447 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-041-886-25

Query Match 55.8%; Score 24; DB 3; Length 1447;  
Best Local Similarity 38.5%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 YXXFLXXRXSP 16  
DB 594 YSLRLAYNRYGP 606

RESULT 11  
PCT-US94-05277-2  
Sequence 2, Application PC/TUS9405277  
GENERAL INFORMATION:  
APPLICANT: Bruskin, Arthur  
APPLICANT: Jarosz, David E.  
APPLICANT: Johnson, Karen  
APPLICANT: Kinzler, Kenneth W.  
APPLICANT: Vogelstein, Bert  
TITLE OF INVENTION: Antibodies Specific for DCC Gene Product  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/05277  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.42709  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202.508.9100  
TELEFAX: 202.508.9299  
TELEX: 197430 BBWB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1447 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-05277-2

Query Match 55.8%; Score 24; DB 5; Length 1447;  
Best Local Similarity 38.5%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 YXXFLXXRXSP 16

DB 594 YSLRLAYNRYGP 606

RESULT 12  
US-08-923-454A-14  
Sequence 14, Application US/08923454A  
Patent No. 6004794  
GENERAL INFORMATION:  
APPLICANT: Creasy, Caretha  
APPLICANT: Livi, George  
APPLICANT: Kattan, Eric  
APPLICANT: Clinkenbeard, Helen  
APPLICANT: Browne, Michael  
APPLICANT: Southan, Christopher  
TITLE OF INVENTION: HUMAN SERINE PROTEASE  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithkline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/923,454A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/025436  
FILING DATE: 06-SEPT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: P50547  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 491 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
US-08-923-454A-14

Query Match 53.5%; Score 23; DB 3; Length 491;  
Best Local Similarity 55.6%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXRXSP 16  
DB 480 FLKTRGSP 488

RESULT 13  
US-08-350-741-3  
Sequence 3, Application US/08350741  
Patent No. 5804194  
GENERAL INFORMATION:  
APPLICANT: DOUGAN G.,  
APPLICANT: CHARLES I.G.,  
APPLICANT: HORMACHE C.E.,

APPLICANT: JOHNSON K.S.,  
APPLICANT: CHATFIELD S.N.  
TITLE OF INVENTION: LIVE VACCINES  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON and VANDERHYE PC  
STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,741  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/952,737  
FILING DATE: 09-MAY-1994  
APPLICATION NUMBER: US 07/952,737  
FILING DATE: 30-NOV-1992  
APPLICATION NUMBER: GB 9007194.5  
FILING DATE: 30-MAR-1990  
APPLICATION NUMBER: PCT/GB91/00484  
FILING DATE: 28-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 117-158  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
TELEX: 200797 NIXN UR  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 492 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-350-741-3

Query Match 53.5%; Score 23; DB 1; Length 492;  
Best Local Similarity 55.6%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 8 FLXXXRXSP 16  
Db 481 FLKTRGSP 489  
RESULT 14  
US-08-463-875A-3  
Sequence 3, Application US/08463875A  
Patent No. 5980907  
GENERAL INFORMATION:  
APPLICANT: DOUGAN, Gordon  
APPLICANT: CHARLES, Ian G.  
APPLICANT: HORMACHE, Carlos E.  
APPLICANT: JOHNSON, Kevin S.  
APPLICANT: CHATFIELD, Steven N.  
TITLE OF INVENTION: LIVE VACCINES  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON and VANDERHYE PC  
STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22201-4714

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,875A  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/340,741  
FILING DATE: 07-DEC-1994  
APPLICATION NUMBER: US 07/952,737  
FILING DATE: 30-NOV-1992  
APPLICATION NUMBER: GB 9007194.5  
FILING DATE: 30-MAR-1990  
APPLICATION NUMBER: PCT/GB91/00484  
FILING DATE: 28-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 117-158  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
TELEX: 200797 NIXN UR  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 492 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-463-875A-3

Query Match 53.5%; Score 23; DB 2; Length 492;  
Best Local Similarity 55.6%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 8 FLXXXRXSP 16  
Db 481 FLKTRGSP 489  
RESULT 15  
US-09-489-039A-8921  
Sequence 8921, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 8921  
LENGTH: 506  
TYPE: PPT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-8921

Query Match 53.5%; Score 23; DB 4; Length 506;  
Best Local Similarity 55.6%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 8 FLXXXRXSP 16  
Db 234 FLTRHFRSP 242

Query Match 53.5%; Score 23; DB 4; Length 506;  
Best Local Similarity 55.6%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 8 FLXXXRXSP 16  
Db 234 FLTRHFRSP 242

Search completed: October 5, 2004, 16:15:40  
Job time : 22.6986 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:12:48 ; Search time 77.1507 Seconds  
(without alignments)  
66.737 Million cell updates/sec

Title: US-09-973-473A-24

Perfect score: 43

Sequence: 1 XXXXXFLXXRXSP 16

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	67.4	208	16	US-10-437-963-111469
2	26	60.5	16	10	US-09-973-473-7
3	26	60.5	16	10	US-09-973-473-16
4	26	60.5	16	10	US-09-973-473-24
5	26	60.5	120	13	US-10-096-703-3
6	26	60.5	120	13	US-10-096-703-4
7	26	60.5	172	16	US-10-767-701-56962
8	25	58.1	16	10	US-09-973-473-6
9	25	58.1	16	10	US-09-973-473-15
10	25	58.1	69	12	US-10-424-599-151268
11	25	58.1	98	12	US-10-424-599-156773
12	25	58.1	117	14	US-10-161-051-157
13	25	58.1	118	14	US-10-353-929-48
14	25	58.1	138	9	US-09-925-301-1164
15	25	58.1	163	16	US-10-437-963-106375

16	25	58.1	453	16	US-10-437-963-204135	Sequence 204135,
17	25	58.1	964	14	US-10-017-161-710	Sequence 710, App
18	25	58.1	964	15	US-10-292-798-622	Sequence 622, App
19	24	55.8	16	10	US-09-973-473-17	Sequence 17, Appl
20	24	55.8	23	9	US-09-984-245-254	Sequence 254, Appl
21	24	55.8	23	10	US-09-966-262-254	Sequence 254, App
22	24	55.8	23	10	US-09-983-966-254	Sequence 254, App
23	24	55.8	23	12	US-10-059-395-254	Sequence 254, App
24	24	55.8	23	14	US-10-143-090-254	Sequence 254, App
25	24	55.8	58	12	US-10-424-599-150803	Sequence 150803,
26	24	55.8	64	9	US-09-984-245-253	Sequence 253, App
27	24	55.8	64	10	US-09-966-262-253	Sequence 253, App
28	24	55.8	64	10	US-09-983-966-253	Sequence 253, App
29	24	55.8	64	12	US-10-059-395-253	Sequence 253, App
30	24	55.8	64	14	US-10-143-090-253	Sequence 253, App
31	24	55.8	66	16	US-10-437-963-106503	Sequence 106503,
32	24	55.8	67	12	US-10-424-599-238176	Sequence 238176,
33	24	55.8	78	16	US-10-437-963-199363	Sequence 199363,
34	24	55.8	100	12	US-10-424-599-182814	Sequence 182814,
35	24	55.8	163	16	US-10-437-963-186461	Sequence 186461,
36	24	55.8	167	12	US-10-425-114-51803	Sequence 51803, A
37	24	55.8	171	12	US-10-425-114-67070	Sequence 67070, A
38	24	55.8	171	12	US-10-425-114-67093	Sequence 67093, A
39	24	55.8	171	12	US-10-425-114-67115	Sequence 67115, A
40	24	55.8	171	12	US-10-425-114-67179	Sequence 67179, A
41	24	55.8	250	16	US-10-437-963-157000	Sequence 157000,
42	24	55.8	335	15	US-10-369-493-10615	Sequence 10615, A
43	24	55.8	355	14	US-10-156-761-9765	Sequence 9765, Ap
44	24	55.8	440	15	US-10-369-493-18063	Sequence 18063, A
45	24	55.8	591	15	US-10-310-154-519	Sequence 519, App

#### ALIGNMENTS

##### RESULT 1

US-10-437-963-111469  
; Sequence 111469, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barzduk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 111469  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(208)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_15445C.1.pep  
US-10-437-963-111469

Query Match 67.4%; Score 29; DB 16; Length 208;  
Best Local Similarity 46.2%; Pred. No. 6.5;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXRXSP 16

DB 192 YGHDFLRSARASP 204

Db	4	YERSFMLSRLROSP	16	
RESULT 2				
US-09-973-473-7				
; Sequence 7, Application US/09973473				
; Publication No. US20030041341A1				
; GENERAL INFORMATION:				
; APPLICANT: SONENBERG, Nahum				
; APPLICANT: TREMBLAY, Michel				
; APPLICANT: TSUKIYAMA-KOHARA, KYOKO				
; TITLE OF INVENTION: NON-HUMAN TRANSGENIC ANIMAL WHOSE GERM CELLS AND				
; TITLE OF INVENTION: SOMATIC CELLS CONTAIN A KNOCKOUT MUTATION IN DNA				
; TITLE OF INVENTION: ENCODING 4E-BP1				
; FILE REFERENCE: 514012000400				
; CURRENT APPLICATION NUMBER: US/09/973,473				
; CURRENT FILING DATE: 2001-10-03				
; PRIOR APPLICATION NUMBER: PCT/CA00/00388				
; PRIOR FILING DATE: 2000-04-07				
; PRIOR APPLICATION NUMBER: 60/128,559				
; PRIOR FILING DATE: 1999-04-09				
; PRIOR APPLICATION NUMBER: 60/179,743				
; PRIOR FILING DATE: 2000-02-02				
; NUMBER OF SEQ ID NOS: 27				
; SOFTWARE: Patentin Ver. 2.1				
; SEQ ID NO 7				
; LENGTH: 16				
; TYPE: PRT				
; ORGANISM: Homo sapiens				
US-09-973-473-7				
Query Match 60.5%; Score 26; DB 10; Length 16;				
Best Local Similarity 46.2%; Pred. No. 4;				
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;				
QY	4	YXXFLXXXRSP	16	
Db	4	YDRKFLDRNSP	16	
RESULT 3				
US-09-973-473-16				
; Sequence 16, Application US/09973473				
; Publication No. US20030041341A1				
; GENERAL INFORMATION:				
; APPLICANT: SONENBERG, Nahum				
; APPLICANT: TREMBLAY, Michel				
; APPLICANT: TSUKIYAMA-KOHARA, KYOKO				
; TITLE OF INVENTION: NON-HUMAN TRANSGENIC ANIMAL WHOSE GERM CELLS AND				
; TITLE OF INVENTION: SOMATIC CELLS CONTAIN A KNOCKOUT MUTATION IN DNA				
; TITLE OF INVENTION: ENCODING 4E-BP1				
; FILE REFERENCE: 514012000400				
; CURRENT APPLICATION NUMBER: US/09/973,473				
; CURRENT FILING DATE: 2001-10-03				
; PRIOR APPLICATION NUMBER: PCT/CA00/00388				
; PRIOR FILING DATE: 2000-04-07				
; PRIOR APPLICATION NUMBER: 60/128,559				
; PRIOR FILING DATE: 1999-04-09				
; PRIOR APPLICATION NUMBER: 60/179,743				
; PRIOR FILING DATE: 2000-02-02				
; NUMBER OF SEQ ID NOS: 27				
; SOFTWARE: Patentin Ver. 2.1				
; SEQ ID NO 16				
; LENGTH: 16				
; TYPE: PRT				
; ORGANISM: Bombyx mori				
US-09-973-473-16				
Query Match 60.5%; Score 26; DB 10; Length 16;				
Best Local Similarity 38.5%; Pred. No. 4;				
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;				
QY	4	YXXFLXXXRSP	16	
Db	4	YXXFLXXXRSP	16	
Query Match 60.5%; Score 26; DB 10; Length 16;				
Best Local Similarity 100.0%; Pred. No. 4;				
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
US-09-973-473-24				
; Sequence 24, Application US/09973473				
; Publication No. US20030041341A1				
; GENERAL INFORMATION:				
; APPLICANT: SONENBERG, Nahum				
; APPLICANT: TREMBLAY, Michel				
; APPLICANT: TSUKIYAMA-KOHARA, KYOKO				
; TITLE OF INVENTION: NON-HUMAN TRANSGENIC ANIMAL WHOSE GERM CELLS AND				
; TITLE OF INVENTION: SOMATIC CELLS CONTAIN A KNOCKOUT MUTATION IN DNA				
; TITLE OF INVENTION: ENCODING 4E-BP1				
; FILE REFERENCE: 514012000400				
; CURRENT APPLICATION NUMBER: US/09/973,473				
; CURRENT FILING DATE: 2001-10-03				
; PRIOR APPLICATION NUMBER: PCT/CA00/00388				
; PRIOR FILING DATE: 2000-04-07				
; PRIOR APPLICATION NUMBER: 60/128,559				
; PRIOR FILING DATE: 1999-04-09				
; PRIOR APPLICATION NUMBER: 60/179,743				
; PRIOR FILING DATE: 2000-02-02				
; NUMBER OF SEQ ID NOS: 27				
; SOFTWARE: Patentin Ver. 2.1				
; SEQ ID NO 24				
; LENGTH: 16				
; TYPE: PRT				
; ORGANISM: Artificial Sequence				
; FEATURE:				
; OTHER INFORMATION: Description of Artificial Sequence: synthetic				
; OTHER INFORMATION: peptide				
; FEATURE:				
; NAME/KEY: VARIANT				
; LOCATION: (1)				
; OTHER INFORMATION: x = positively charged amino acid				
; FEATURE:				
; NAME/KEY: VARIANT				
; LOCATION: (6)				
; OTHER INFORMATION: x = positively charged amino acid				
; FEATURE:				
; NAME/KEY: VARIANT				
; LOCATION: (2)..(3)				
; OTHER INFORMATION: x = hydrophobic amino acid				
; FEATURE:				
; NAME/KEY: VARIANT				
; LOCATION: (10)				
; OTHER INFORMATION: x = hydrophobic amino acid				
; FEATURE:				
; NAME/KEY: VARIANT				
; LOCATION: (5)				
; OTHER INFORMATION: x = negatively charged amino acid				
; FEATURE:				
; NAME/KEY: VARIANT				
; LOCATION: (7)				
; OTHER INFORMATION: x - any amino acid				
; FEATURE:				
; NAME/KEY: VARIANT				
; LOCATION: (11)..(12)				
; OTHER INFORMATION: x = any amino acid				
; FEATURE:				
; NAME/KEY: VARIANT				
; LOCATION: (14)				
; OTHER INFORMATION: x = any amino acid				
US-09-973-473-24				
Query Match 60.5%; Score 26; DB 10; Length 16;				
Best Local Similarity 100.0%; Pred. No. 4;				
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	4	YXXFLXXXRSP	16	
Db	4	YXXFLXXXRSP	16	



RESULT 5  
US-10-096-703-3  
; Sequence 3, Application US/10096703  
; Publication No. US20020132330A1  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; ; Hawkins, Phillip R.  
; TITLE OF INVENTION: NEW TRANSLATIONAL REGULATOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/096,703  
; FILING DATE: 12-Mar-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/215,063  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0310 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 561632  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-10-096-703-3  
Query Match 60.5%; Score 26; DB 13; Length 120;  
Best Local Similarity 46.2%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Oy 4 YXXFLXXXRSP 16  
Db 54 YDRKFLDRRNSP 66  
RESULT 6  
US-10-096-703-4  
; Sequence 4, Application US/10096703  
; Publication No. US20020132330A1  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; ; Hawkins, Phillip R.  
; TITLE OF INVENTION: NEW TRANSLATIONAL REGULATOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA

COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/096,703  
FILING DATE: 12-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/215,063  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0310 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 120 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1658516  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-096-703-4  
Query Match 60.5%; Score 26; DB 13; Length 120;  
Best Local Similarity 46.2%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Oy 4 YXXFLXXXRSP 16  
Db 54 YDRKFLDRRNSP 66  
RESULT 7  
US-10-767-701-56962  
; Sequence 56962, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 56962  
; LENGTH: 172  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(172)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 30951574.pep  
US-10-767-701-56962  
Query Match 60.5%; Score 26; DB 16; Length 172;  
Best Local Similarity 38.5%; Pred. No. 38;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXRXSP 16  
| | | | |  
Db 5 YAAFLGTVEAP 17

## RESULT 8

US-09-973-473-6  
; Sequence 6, Application US/09973473  
; Publication No. US20030041341A1  
; GENERAL INFORMATION:  
; APPLICANT: SONENBERG, Nahum  
; APPLICANT: TREMBLAY, Michel  
; APPLICANT: TSUKIYAMA-KOHARA, KYOKO  
; TITLE OF INVENTION: NON-HUMAN TRANSGENIC ANIMAL WHOSE GERM CELLS AND  
; TITLE OF INVENTION: SOMATIC CELLS CONTAIN A KNOCKOUT MUTATION IN DNA  
; FILE OF INVENTION: ENCODING 4E-BP1  
; FILE REFERENCE: 51401200400  
; CURRENT APPLICATION NUMBER: US/09/973,473  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: PCT/CA00/00388  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: 60/128,559  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 60/179,743  
; PRIOR FILING DATE: 2000-02-02  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-973-473-6

Query Match 58.1%; Score 25; DB 10; Length 16;  
Best Local Similarity 46.2%; Pred. No. 7.6;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXRXSP 16  
| | | | |  
Db 4 YDRKFLMECRNSP 16

## RESULT 9

US-09-973-473-15  
; Sequence 15, Application US/09973473  
; Publication No. US20030041341A1  
; GENERAL INFORMATION:  
; APPLICANT: SONENBERG, Nahum  
; APPLICANT: TREMBLAY, Michel  
; APPLICANT: TSUKIYAMA-KOHARA, KYOKO  
; TITLE OF INVENTION: NON-HUMAN TRANSGENIC ANIMAL WHOSE GERM CELLS AND  
; TITLE OF INVENTION: SOMATIC CELLS CONTAIN A KNOCKOUT MUTATION IN DNA  
; FILE OF INVENTION: ENCODING 4E-BP1  
; FILE REFERENCE: 51401200400  
; CURRENT APPLICATION NUMBER: US/09/973,473  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: PCT/CA00/00388  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: 60/128,559  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: 60/179,743  
; PRIOR FILING DATE: 2000-02-02  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Halocynthia roretzi  
US-09-973-473-15

Query Match 58.1%; Score 25; DB 10; Length 16;  
Best Local Similarity 46.2%; Pred. No. 7.6;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXRXSP 16  
| | | | |  
Db 4 YDRFLKCRDSP 16

## RESULT 10

US-10-424-599-151268  
; Sequence 151268, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: KOVALIC David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 151268  
; LENGTH: 69  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_107618C.1.pep  
US-10-424-599-151268

Query Match 58.1%; Score 25; DB 12; Length 69;  
Best Local Similarity 55.6%; Pred. No. 31;  
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXXRXSP 16  
| | | | |  
Db 13 FLTRRSSP 21

## RESULT 11

US-10-424-599-156773  
; Sequence 156773, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: KOVALIC David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 156773  
; LENGTH: 98  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_112589C.1.pep  
US-10-424-599-156773

Query Match 58.1%; Score 25; DB 12; Length 98;  
Best Local Similarity 41.7%; Pred. No. 43;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXRXS 15  
| | | | |  
Db 14 YSSAFLSASRRS 25

## RESULT 12

US-10-161-051-157  
; Sequence 157, Application US/10161051

```

; Publication No. US20030152945A1
; GENERAL INFORMATION:
; APPLICANT: Peter Deak
; APPLICANT: David M Glover
; APPLICANT: Carol Midgley
; TITLE OF INVENTION: Cell cycle progression proteins
; FILE REFERENCE: CCI-021CP
; CURRENT APPLICATION NUMBER: US/10/161,051
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: GB 0007268.6
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 157
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-161-051-157

Query Match 58.1%; Score 25; DB 14; Length 117;
Best Local Similarity 38.5%; Pred. No. 51;
Matches 5; Conservative 1; Mismatches 7; Indels

QY 4 YXXFLLXXXRSP 16
| | | |
DB 54 YERAFMKNLRGSP 66

RESULT 13
US-10-353-929-48
; Sequence 48, Application US/10353929
; Publication No. US20030175288A1
; GENERAL INFORMATION:
; APPLICANT: ITOH, Kyogo
; TITLE OF INVENTION: Tumor antigen
; FILE REFERENCE: GP01-1024
; CURRENT APPLICATION NUMBER: US/10/353,929
; CURRENT FILING DATE: 2003-01-30
; PRIOR APPLICATION NUMBER: JP P2000-231814
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-353-929-48

Query Match 58.1%; Score 25; DB 14; Length 118;
Best Local Similarity 46.2%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 7; Indels

QY 4 YXXFLLXXXRSP 16
| | | |
DB 54 YDRKFLMECRNSP 66

RESULT 14
US-09-925-301-1164
; Sequence 1164, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:58:01 ; Search time 19.5068 Seconds  
(without alignments)  
78.899 Million cell updates/sec

Title: US-09-973-473A-24  
Perfect score: 43  
Sequence: 1 XXXYXXFLXXRXSP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 28366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

\*Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	62.8	436	2 E71493	probable RNA polym
2	27	62.8	721	2 A87128	conserved hypothet
3	26	60.5	120	2 S50867	4E-BP2 protein - h
4	25	58.1	109	2 G71013	hypothetical prote
5	25	58.1	117	2 A55258	insulin-stimulated
6	25	58.1	117	2 A57396	PHAS-I protein - m
7	25	58.1	118	2 S50866	translation initia
8	25	58.1	547	2 A95861	hypothetical prote
9	25	58.1	1562	2 S33069	probable membrane
10	24	55.8	324	2 A72285	conserved hypothet
11	24	55.8	355	2 T34711	uroporphyrinogen d
12	24	55.8	383	2 F75426	sensor histidine k
13	24	55.8	530	2 A36276	glucuronosyltransf
14	24	55.8	541	2 T23689	hypothetical prote
15	24	55.8	668	2 C86476	protein F1504.45 [
16	24	55.8	711	2 D86296	hypothetical prote
17	24	55.8	1447	2 A54100	tumor suppressor p
18	23	53.5	114	2 J55860	polyketide synthas
19	23	53.5	164	2 F75526	hypothetical prote
20	23	53.5	206	2 T16820	hypothetical prote
21	23	53.5	284	2 S5167	probable membrane
22	23	53.5	313	2 AB3052	hypothetical prote
23	23	53.5	313	2 B98234	hypothetical prote
24	23	53.5	315	2 A95275	probable ABC trans
25	23	53.5	395	2 AC3571	acyl-CoA dehydroge
26	23	53.5	530	2 S68200	glucuronosyltransf
27	23	53.5	530	2 S00163	glucuronosyltransf
28	23	53.5	530	2 S07390	glucuronosyltransf
29	23	53.5	574	2 T41068	hypothetical prote

30	23	53.5	534	1 WLWL51	E1 protein - human
31	23	53.5	730	2 H86295	hypothetical prote
32	23	53.5	762	2 B70838	hypothetical prote
33	23	53.5	798	2 S40052	glycogen phosphory
34	23	53.5	805	2 A10176	probable oxidoredu
35	23	53.5	827	2 S48465	6-phosphofructo-2-
36	23	53.5	1004	2 A39611	probable GTP-bindi
37	23	53.5	1374	2 T30809	plasmaenogen relate
38	23	53.5	1628	2 T38055	hypothetical prote
39	23	53.5	1874	1 JQ0533	genome polyprotein
40	23	53.5	4302	2 A38971	polycystic kidney
41	22	51.2	104	2 S56778	probable membrane
42	22	51.2	134	2 G96499	hypothetical prote
43	22	51.2	161	2 T48932	adenosine receptor
44	22	51.2	171	2 B87018	probable membrane
45	22	51.2	184	2 S41540	hypothetical 21.6K

ALIGNMENTS

RESULT 1

E71493

probable RNA polymerase sigma-54 - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C;Species: Chlamydia trachomatis

C;Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 17-Mar-2000

C;Accession: E71493

R;Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,

<Science 282, 754-759, 1998

A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract

A;Reference number: A71570; MUID:9900809; PMID:9784136

A;Accession: E71493

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-436 <ARN>

A;Cross-references: GB:AE001331; GB:AE001273; NID:g3323046; PIDN:AAC68212.1; PID:g33239054

A;Experimental source: serotype D, strain UW-3/Cx

C;Genetics:

A;Gene: rpon

C;Superfamily: Pseudomonas transcription initiation factor sigma

Query Match 62.8%; Score 27; DB 2; Length 436;

Best Local Similarity 46.2%; Pred. No. 3.3;

Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 4 YXXYFLXXRXSP 16

Db 298 YQEEFLKRTSP 310

RESULT 2

A87128

conserved hypothetical protein ML1751 [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: A87128

R;Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,

cam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

A;Title: Massive gene decay in the leprosy bacillus.

A;Reference number: A86909; MUID:21128732; PMID:11234002

A;Accession: A87128

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-721 <STO>

A;Cross-references: GB:AL450380; NID:g13093490; PIDN:CAC30704.1; GSPDB:GN00147

C;Genetics:

A;Gene: ML1751

Query Match 62.8%; Score 27; DB 2; Length 721;

Best Local Similarity 46.2%; Pred. No. 5.5;

Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;									
QY	4	YXXFLXXRXSP 16							
Db	281	YILFLEAERLSP 293							
RESULT 3									
4E-BP2 protein - human									
C:Species: Homo sapiens (man)									
C>Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 25-Sep-1998									
C:Accession: S50867									
R:Pause, A.; Belsham, G.J.; Gingras, A.C.; Donze, O.; Lin, T.A.; Lawrence Jr., J.C.; Son									
Nature 371, 762-767, 1994									
A:Title: Insulin-dependent stimulation of protein synthesis by phosphorylation of a regu									
A:Reference number: S50866; MUID:95021760; PMID:7935836									
A:Accession: S50867									
A:Status: preliminary									
A:Molecule type: mRNA									
A:Residues: 1-120 <PAU>									
A:Cross-references: EMBL:U36056; NID:g561631; PID:g561632									
C:Genetics:									
A:Gene: GDB:EIF4EBP2									
A:Cross-references: GDB:437248									
Query Match 60.5%; Score 26; DB 2; Length 120;									
Best Local Similarity 46.2%; Pred. No. 1.9;									
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;									
QY	4	YXXFLXXRXSP 16							
Db	54	YDRKFLDRNSP 66							
RESULT 4									
G71013									
hypothetical protein PH1405 - Pyrococcus horikoshii									
C:Species: Pyrococcus horikoshii									
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000									
C:Accession: G71013									
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki									
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi									
DNA Res. 5, 55-76, 1998									
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a									
A:Reference number: A71000; MUID:98344137; PMID:9679194									
A:Accession: G71013									
A:Status: preliminary; nucleic acid sequence not shown; translation not shown									
A:Molecule type: DNA									
A:Residues: 1-109 <KAW>									
A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30511.1; PID:g3257828									
A:Experimental source: strain OT3									
A>Note: This accession replaces an interim accession for a sequence replaced by GenBank									
C:Genetics:									
A:Gene: PH1405									
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1405									
Query Match 58.1%; Score 25; DB 2; Length 109;									
Best Local Similarity 55.6%; Pred. No. 3.5;									
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;									
QY	8	FLXXRXSP 16							
Db	17	FLASIRTSP 25							
RESULT 5									
A55258									
insulin-stimulated phosphoprotein PHAS-I - rat									
C:Species: Rattus norvegicus (Norway rat)									
C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 08-Oct-1999									
C:Accession: A55258; A54719									
R:Hu, C.; Pang, S.; Kong, X.; Vellica, M.; Lawrence Jr., J.C.									

Proc. Natl. Acad. Sci. U.S.A. 91, 3730-3734, 1994									
A:Title: Molecular cloning and tissue distribution of PHAS-I, an intracellular target for									
A:Reference number: A55258; MUID:94224815; PMID:8170978									
A:Accession: A55258									
A:Molecule type: mRNA									
A:Residues: 1-117 <HUA>									
A:Cross-references: GB:U05014; NID:g468023; PIDN:AAA86938.1; PID:g468024									
R:Haystead, T.A.J.; Haystead, C.M.M.; Hu, C.; Lin, T.A.; Lawrence Jr., J.C.									
J. Biol. Chem. 269, 23185-23191, 1994									
A:Title: Phosphorylation of PHAS-I by mitogen-activated protein (MAP) kinase. Identifica									
A:Reference number: A54719; MUID:94365019; PMID:8083223									
A:Accession: A54719									
A:Molecule type: protein									
A:Residues: 58-69 <HAY>									
C:Keywords: phosphoprotein									
F:64/Binding site: phosphate (Ser) (covalent) (by MAP kinase) #status experimental									
Query Match 58.1%; Score 25; DB 2; Length 117;									
Best Local Similarity 46.2%; Pred. No. 3.7;									
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;									
Qy	4	YXXYFLXXRXSP 16							
Db	53	YDRKFLMECRNSP 65							
RESULT 6									
A57396									
PHAS-I protein - mouse									
C:Species: Mus musculus (house mouse)									
C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 08-Oct-1999									
C:Accession: A57396									
R:Lin, T.A.; Kong, X.; Saltiel, A.R.; Blackshear, P.J.; Lawrence Jr., J.C.									
J. Biol. Chem. 270, 18531-18538, 1995									
A:Title: Control of PHAS-I by insulin in 3T3-L1 adipocytes. Synthesis, degradation, and f									
A:Reference number: A57396; MUID:95355483; PMID:7629182									
A:Accession: A57396									
A:Status: preliminary; not compared with conceptual translation									
A:Molecule type: mRNA									
A:Residues: 1-117 <LIN>									
A:Cross-references: GB:U28656; NID:g881557; PIDN:AAA88818.1; PID:g881558									
Query Match 58.1%; Score 25; DB 2; Length 117;									
Best Local Similarity 46.2%; Pred. No. 3.7;									
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;									
Qy	4	YXXYFLXXRXSP 16							
Db	53	YDRKFLMECRNSP 65							
RESULT 7									
S50866									
translation initiation factor 4E-binding protein 1 - human									
N:Alternate names: 4E-BP1 protein									
C:Species: Homo sapiens (man)									
C>Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 31-Mar-2000									
C:Accession: S50866; JC5899									
R:Pause, A.; Belsham, G.J.; Gingras, A.C.; Donze, O.; Lin, T.A.; Lawrence Jr., J.C.; Son									
Nature 371, 762-767, 1994									
A:Title: Insulin-dependent stimulation of protein synthesis by phosphorylation of a regul									
A:Reference number: S50866; MUID:95021760; PMID:7935836									
A:Accession: S50866									
A:Status: preliminary									
A:Molecule type: mRNA									
A:Residues: 1-118 <PAU>									
A:Cross-references: EMBL:U36055; NID:g561629; PIDN:AAA62269.1; PID:g561630									
R:Nishi, N.; Morino, S.; Tomco, K.; Youtani, T.; Ishida, T.									
J. Biochem. 123, 157-161, 1998									
A:Title: Expression of a synthetic gene for initiation factor 4E-binding protein 1 in Es									
A:Reference number: JC5899; MUID:98162949; PMID:9504423									
A:Accession: JC5899									
A:Molecule type: DNA									

A;Residues: 1-118 <NIS>  
C;Comment: This protein exhibits a reduced level of complex formation with initiation factor 4E.  
C;Genetics:  
A;Gene: GDB:EIF4EBP1  
A;Cross-references: GDB:437247  
C;Keywords: phosphoprotein  
F;65/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 58.1%; Score 25; DB 2; Length 118;  
Best Local Similarity 46.2%; Pred. No. 3.7;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16  
DB 54 YDRKFLMECRNSP 66

RESULT 8  
A5861  
Hypothetical protein Smb20153 [imported] - Sinorhizobium meliloti (strain 1021) megaplas-  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Accession: A95861  
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan-  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A;Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo-  
A;Reference number: A95842; MUID:21396508; PMID:11481431  
A;Accession: A95861  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-547 <KUR>  
A;Cross-references: GB:AL591985; PIDN:CAC48553.1; PID:g15140025; GSPDB:GN00167  
A;Experimental source: strain 1021, megaplasmid pSymb  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
A;Contents: annotation  
C;Genetics:  
A;Gene: Smb20153  
A;Genome: plasmid

Query Match 58.1%; Score 25; DB 2; Length 547;  
Best Local Similarity 38.5%; Pred. No. 17; Mismatches 1; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16  
DB 261 YCAEFMELARLSP 273

RESULT 9  
S53069  
probable membrane protein YMR247c - Yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein YMR9408.09c; hypothetical protein YMR9920.01c  
C;Species: Saccharomyces cerevisiae  
C;Date: 08-Jul-1995 #sequence\_revision 09-Mar-1996 #text\_change 02-Sep-2000  
C;Accession: S53069; S56061  
R;Hunt, S.; Bowman, S.  
submitted to the EMBL Data Library, March 1995  
A;Reference number: S53069  
A;Accession: S53069  
A;Molecule type: DNA  
A;Residues: 1-956 <HUN>  
A;Cross-references: EMBL:248639; NID:g732924; PID:g732925; MIPS:YMR247c  
R;Gentles, S.; Bowman, S.  
submitted to the EMBL Data Library, March 1995  
A;Reference number: S56053  
A;Accession: S56061

A;Molecule type: DNA  
A;Residues: 950-1562 <GEN>  
A;Cross-references: EMBL:Z48756; NID:g736304; PID:g736313; MIPS:YMR247c  
C;Genetics:  
A;Map position: 13R  
C;Superfamily: RING finger homology  
F;208-224/Domain: transmembrane protein  
F;303-319/Domain: transmembrane #status predicted <TM1>  
F;586-602/Domain: transmembrane #status predicted <TM2>  
F;1504-1560/Domain: RING finger homology <RRN>

Query Match 58.1%; Score 25; DB 2; Length 1562;  
Best Local Similarity 55.6%; Pred. No. 50;  
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXXRXSP 16  
DB 342 FLSVSRTPSP 350

RESULT 10  
A72285  
Conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: A72285  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq-  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: A72285  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-324 <ARN>  
A;Cross-references: GB:AB001775; GB:AE000512; NID:g4981732; PIDN:AAD36262.1; PID:g498173;  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1187  
C;Superfamily: Bacillus subtilis conserved hypothetical protein yuef

Query Match 55.8%; Score 24; DB 2; Length 324;  
Best Local Similarity 38.5%; Pred. No. 21;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16  
DB 13 YLVFLVLAKLSP 25

RESULT 11  
T34711  
uroporphyrinogen decarboxylase - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
C;Accession: T34711  
R;Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1998  
A;Reference number: Z21554  
A;Accession: T34711  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-355 <OLI>  
A;Cross-references: EMBL:AL023702; PIDN:CAA19243.1; GSPDB:GN00070; SCOEDB:SC1C3.19  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: hemE; SCOEDB:SC1C3.19  
C;Superfamily: uroporphyrinogen decarboxylase

Query Match 55.8%; Score 24; DB 2; Length 355;  
Best Local Similarity 38.5%; Pred. No. 23;

Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16  
| | | | |  
Db 16 YDSAFLKACREP 28

## RESULT 12

F75426  
sensor histidine kinase/response regulator - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Apr-2000  
C:Accession: F75426  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: F75426

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-383 <WHI>

A:Cross-references: GB:AE001967; GB:AE000513; NID:g6458915; PIDN:AAF10748.1; PID:g645891

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1175

A:Map position: 1

C:Superfamily: sensory transduction system regulatory protein; response regulator homolog

Query Match 55.8%; Score 24; DB 2; Length 383;

Best Local Similarity 55.6%; Pred. No. 25;

Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXXRXSP 16  
| | | | |

Db 36 FLTALRESP 44

## RESULT 13

A36276  
glucuronosyltransferase (EC 2.4.1.17) 5 precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 18-Jan-1991 #sequence\_revision 13-Jan-1993 #text\_change 29-Sep-1999

C:Accession: A36276

R:Mackenzie, P.I.

J. Biol. Chem. 265, 8699-8703, 1990

A:Title: The cDNA sequence and expression of a variant 17beta-hydroxysteroid UDP-glucur

A:Reference number: A36276; MUID:90256795; PMID:1692835

A:Accession: A36276

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-530 <MAC>

A:Cross-references: GB:M3746; GB:J05440; NID:g207570; PIDN:AAA03217.1; PID:g207572

A>Note: the authors translated the codon GAG for residue 530 as Lys

C:Superfamily: glucuronosyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 55.8%; Score 24; DB 2; Length 530;

Best Local Similarity 55.6%; Pred. No. 35;

Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXXRXSP 16  
| | | | |

Db 169 FLYSLRASP 177

## RESULT 14

T23689  
hypothetical protein M03C11.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T23689

R:McMurray, A.

submitted to the EMBL Data Library, April 1995

A:Reference number: Z19783

A:Accession: T23689

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-541 <WIL>

A:Cross-references: EMBL:Z49128; PIDN:CAA8954.1; GSPDB:GN00021; CESP:M03C11.4

A:Experimental source: clone M03C11

C:Genetics:

A:Gene: CESP:M03C11.4

A:Map position: 3

A:Introns: 118/3; 158/1; 298/2; 318/2

Query Match 55.8%; Score 24; DB 2; Length 541;

Best Local Similarity 38.5%; Pred. No. 35;

Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16  
| | | | |

Db 373 FMESFLDLRASP 385

## RESULT 15

C86476

protein F1504.45 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C86476

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C86476

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-668 <STO>

A:Cross-references: GB:AE005172; NID:g8778357; PIDN:AAF79365.1; GSPDB:GN00141

C:Genetics:

A:Gene: F1504.45

A:Map position: 1

Query Match 55.8%; Score 24; DB 2; Length 668;

Best Local Similarity 30.8%; Pred. No. 44;

Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16  
| | | | |

Db 346 YLASPIRAVRVAP 358

Search completed: October 5, 2004, 16:13:57

Job time : 21.5068 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:36 ; Search time 11.1781 Seconds  
(without alignments)  
74.532 Million cell updates/sec

Title: US-09-973-473a-24

Perfect score: 43

Sequence: 1 XXXYXXFLXXXRXSP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	60.5	120	1	4BP2_HUMAN
2	26	60.5	120	1	4BP2_MOUSE
3	25	58.1	117	1	4BP1_MOUSE
4	25	58.1	117	1	4BP1_RAT
5	25	58.1	118	1	4BP1_HUMAN
6	25	58.1	1562	1	YH81_YEAST
7	24	55.8	321	1	CXA9_BOVIN
8	24	55.8	321	1	CXA9_HUMAN
9	24	55.8	324	1	YH87_THEMA
10	24	55.8	355	1	DCUP_STRAW
11	24	55.8	355	1	DCUP_STROO
12	24	55.8	530	1	UDB6_RAT
13	24	55.8	551	1	AGLA_RHINE
14	24	55.8	1447	1	DCU_HUMAN
15	24	55.8	1447	1	DCU_MOUSE
16	24	55.8	4513	1	DV1B_CHLRE
17	23	53.5	165	1	LSPA_CHLCV
18	23	53.5	258	1	ATE_BRAJA
19	23	53.5	530	1	UDB3_RAT
20	23	53.5	530	1	UBB5_MOUSE
21	23	53.5	530	1	UBB5_RAT
22	23	53.5	533	1	LEU1_MICAE
23	23	53.5	560	1	EAA5_HUMAN
24	23	53.5	634	1	VEL1_HPV51
25	23	53.5	777	1	CUL3_CAEEL
26	23	53.5	798	1	PHSG_BACSU
27	23	53.5	827	1	6P21_YEAST
28	23	53.5	1003	1	MV10_HUMAN
29	23	53.5	1004	1	MV10_MOUSE
30	23	53.5	1073	1	UBA4_HUMAN
31	23	53.5	1628	1	YATE_SCHPO
32	23	53.5	1874	1	POLR_KYMWJ
33	23	53.5	4303	1	PKD1_HUMAN

#### ALIGNMENTS

RESULT 1  
4BP2\_HUMAN  
ID 4BP2\_HUMAN STANDARD; PRT; 120 AA.  
AC Q13542;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Eukaryotic translation initiation factor 4E binding protein 2 (4E-BP2) (eIF4E-binding protein 2).  
GN EIF4EBP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND INTERACTION WITH EIF4E.  
RC TISSUE=Placenta;  
RX MEDLINE=95021760; PubMed=7935836;  
RA Pause A., Belsham G.J., Gingras A.-C., Donze O., Lin T.-A., Lawrence J.C. Jr., Sonenberg N.;  
RA "Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function.";  
RL Nature 371:762-767(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kainine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S., Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y., Phelan M., Farmer A.;  
RT "Cloning of human full-length CDSs in BD Creator(TM) system donor vector.";  
RL Submitted (MAY-2003) to the ENBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
TISSUE=Lung, and Uterus;  
MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: Regulates eIF4E activity by preventing its assembly into

34 22 51.2 100 1 4BP3\_HUMAN O60516 homo sapien  
35 22 51.2 101 1 4BP3\_MOUSE Q80VV3 mus musculus  
36 22 51.2 104 1 YJA7\_YEAST P47080 saccharomyc  
37 22 51.2 184 1 YARQ\_ACTPL P46393 actinobacil  
38 22 51.2 236 1 CEMA\_MESVI Q9mun7 mesostigma  
39 22 51.2 265 1 COQ4\_HUMAN Q9V3A0 homo sapien  
40 22 51.2 267 1 DNAL\_SCHPO Q10322 schizosach  
41 22 51.2 296 1 PROB\_TREPA P74936 treponema p  
42 22 51.2 326 1 MENC\_MYCTU O06419 mycobacteri  
43 22 51.2 329 1 DHQA\_EMENI P25415 emericella  
44 22 51.2 372 1 YOIG\_BACSU P54524 bacillus su  
45 22 51.2 396 1 O49A\_DROME Q9V6A9 drosophila

CC the EIF4F complex. Mediates the regulation of protein translation  
 CC by hormones, growth factors and other stimuli that signal through  
 CC the MAP kinase pathway.  
 CC -!- SUBUNIT: Nonphosphorylated EIF4EBP2 interacts with EIF4E.  
 CC -!- PTM: Phosphorylated on serine and threonine residues in response  
 CC to insulin, EGF and PDGF.  
 CC -!- SIMILARITY: Belongs to the EIF4E binding protein family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L36056; AA62270.1; -.  
 CC DR EMBL; BT007317; AAP35981.1; -.  
 CC DR EMBL; BC05057; AAH05057.1; -.  
 CC DR EMBL; BC050633; AAH50633.1; -.  
 CC DR PIR; S50867; S50867  
 CC DR Genew; HGNC:1289; EIF4EBP2.  
 CC DR MIM; 60224; -.  
 CC DR GO; GO:006445; P:regulation of translation; TAS.  
 CC DR InterPro; IPR008606; EIF4EBP.  
 CC DR Pfam; PF05456; EIF4EBP; 1.  
 CC KW Translation regulation; Protein synthesis inhibitor; Phosphorylation.  
 CC  
 CC FT MOD\_RES 1 1 BLOCKED (PROBABLE).  
 CC FT MOD\_RES 65 65 PHOSPHORYLATION (BY MAPK1 AND MAPK3) (BY  
 CC SIMILARITY).  
 CC FT SEQUENCE 120 AA; 12939 MW; B8F109261A504193 CRC64;  
 CC  
 CC Query Match 60.5%; Score 26; DB 1; Length 120;  
 CC Best Local Similarity 46.2%; Pred. NO. 0.77;  
 CC Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 CC  
 CC QY 4 YXXFLXXXXXSP 16  
 CC Db 54 YDRKFLDNRNSP 66  
 CC  
 CC RESULT 2  
 CC ID\_4BP2\_MOUSE STANDARD; PRT; 120 AA.  
 CC AC P704J5;  
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 CC DE Eukaryotic translation initiation factor 4E binding protein 2 (4E-BP2)  
 CC DE (EIF4E-binding protein 2) (Phosphorylated heat- and acid-stable  
 CC DE protein regulated by insulin 2) (PHAS-II).  
 CC GN EIF4EBP2.  
 CC OS Mus musculus (Mouse).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC OX NCBI\_TaxID=10090;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=97094737; PubMed=8939971;  
 CC RA Lin T.A., Lawrence J.C. Jr.;  
 CC RT "Control of the translational regulators PHAS-I and PHAS-II by insulin  
 CC and cAMP in 3T3-L1 adipocytes.";  
 CC RL J. Biol. Chem. 271:30199-30204 (1996).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE=Colon;  
 CC RX MEDLINE=22388257; PubMed=12477932;  
 CC RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 CC RA Klausner R.D., Colling F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 CC RA Atschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 CC RA Hopkins R.F., Jordan H., Moore S.I., Max S.I., Wang J., Hsieh F.,  
 CC RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 CC RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC -!- FUNCTION: Regulates EIF4E activity by preventing its assembly into  
 CC the EIF4F complex. Mediates the regulation of protein translation  
 CC by hormones, growth factors and other stimuli that signal through  
 CC the MAP kinase pathway (By similarity).  
 CC -!- SUBUNIT: Nonphosphorylated EIF4EBP2 interacts with EIF4E (By  
 CC similarity).  
 CC -!- PTM: Phosphorylated on serine and threonine residues in response  
 CC to insulin, EGF and PDGF (By similarity).  
 CC -!- SIMILARITY: Belongs to the EIF4E binding protein family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U75530; AAC52899.1; -.  
 CC DR EMBL; BC015082; AAH15082.1; -.  
 CC DR MGD; MGI:109198; Eif4ebp2.  
 CC DR GO; GO:0005515; P:protein binding; IPI.  
 CC DR GO; GO:0019933; P:cAMP-mediated signaling; IDA.  
 CC DR GO; GO:0008286; P:insulin receptor signaling pathway; IDA.  
 CC DR InterPro; IPR008606; EIF4EBP.  
 CC DR Pfam; PF05456; EIF4EBP; 1.  
 CC KW Translation regulation; Protein synthesis inhibitor; Phosphorylation.  
 CC  
 CC FT MOD\_RES 1 1 BLOCKED (PROBABLE).  
 CC FT MOD\_RES 65 65 PHOSPHORYLATION (BY MAPK1 AND MAPK3) (BY  
 CC SIMILARITY).  
 CC FT SEQUENCE 120 AA; 12898 MW; 0A1ACC082583F769 CRC64;  
 CC  
 CC Query Match 60.5%; Score 26; DB 1; Length 120;  
 CC Best Local Similarity 46.2%; Pred. NO. 0.77;  
 CC Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 CC  
 CC QY 4 YXXFLXXXXXSP 16  
 CC Db 54 YDRKFLDNRNSP 66  
 CC  
 CC RESULT 3  
 CC ID\_4BP1\_MOUSE STANDARD; PRT; 117 AA.  
 CC AC Q60876; Q9CZ40;  
 CC DT 10-OCT-2003 (Rel. 42, Created)  
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 CC DE Eukaryotic translation initiation factor 4E binding protein 1 (4E-BP1)  
 CC DE (EIF4E-binding protein 1) (Phosphorylated heat- and acid-stable  
 CC DE protein regulated by insulin 1) (PHAS-I).  
 CC GN EIF4EBP1.  
 CC OS Mus musculus (Mouse).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC OX NCBI\_TaxID=10090;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.; FUNCTION, TISSUE SPECIFICITY, PHOSPHORYLATION, AND  
 CC INTERACTION WITH EIF4E.

RX MEDLINE=95355483; PubMed=7629182;  
RA Lin T.-A., Kong X., Sattiel A.R., Blackhear P.J., Lawrence J.C. Jr.,  
RT "Control of PHAS-I by insulin in 3T3-L1 adipocytes. Synthesis,  
RT degradation, and phosphorylation by a rapamycin-sensitive and mitogen-  
RT activated protein kinase-independent pathway".  
RL J. Biol. Chem. 270:18531-18538(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=22354683; PubMed=12466851;  
RA Okazaki Y., Furuno M., Kaeukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schrim L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Guetinch S., Hirokawa N., Jackson I.J., Jarvis B.D.,  
RA Kanai A., Kawai H., Kawaawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perteau G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yangisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs".  
RL Nature 420:563-573(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Mammary gland;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.B.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP FUNCTION: Regulates eIF4E activity by preventing its assembly into  
CC the eIF4F complex. Mediates the regulation of protein translation  
CC by hormones, growth factors and other stimuli that signal through  
CC the MAP kinase pathway.  
CC -1- SUBUNIT: Nonphosphorylated eIF4EBP1 competes with eIF4G to  
CC interact with eIF4E; insulin stimulated MAP-kinase (MAPK1 and  
CC MAPK3) phosphorylation of eIF4EBP1 causes dissociation of the  
CC complex allowing eIF4G to bind and consequent initiation of  
CC translation. Rapamycin can attenuate insulin stimulation, mediated  
CC by FKBP5.

CC -1- TISSUE SPECIFICITY: Highest expression in fat cells.  
CC -1- PTM: Phosphorylated on serine and threonine residues in response  
CC to insulin, EGF and PDGF.  
CC -1- SIMILARITY: Belongs to the eIF4E binding protein family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U28656; AAA88818.1; -;  
DR EMBL; AK013033; BAB28612.1; -;  
DR EMBL; BC002045; AA02045.1; -;  
DR PIR; A57396; A57396;  
DR MGD; MGI:103267; Eif4ebp1.  
DR GO; GO:0005515; F:protein binding; IPI.  
DR GO; GO:0008286; P:insulin receptor signaling pathway; IDA.  
DR InterPro; IPR008606; EIF4EBP.  
DR Pfam; PF05456; EIF4EBP; 1.  
DR Translation regulation; Protein synthesis inhibitor; Phosphorylation.  
FT MOD\_RES 1 1 BLOCKED (PROBABLE).  
FT MOD\_RES 64 64 PHOSPHORYLATION (BY MAPK1 AND MAPK3) (BY  
FT SIMILARITY).  
FT CONFLICT 93 93 S -> N (IN REF. 2).  
SQ SEQUENCE 117 AA; 12325 MW; 3458D5687468A7EA CRC64;  
  
Query Match 58.1%; Score 25; DB 1; Length 117;  
Best Local Similarity 46.2%; Pred. No. 1.6;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 4 YXXFLXXXRXSP 16  
Db 53 YDRFLMECRNSP 65  
  
RESULT 4  
4BP1 RAT STANDARD; PRT; 117 AA.  
AC Q62622;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Eukaryotic translation initiation factor 4E binding protein 1 (4E-BP1)  
DE (eIF4E-binding protein 1) (Phosphorylated heat- and acid-stable  
DE protein regulated by insulin 1) (PHAS-I).  
DE EIF4EBP1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OC NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., SEQUENCE OF 18-33; 43-53; 62-80 AND 98-117,  
RP PHOSPHORYLATION, AND TISSUE SPECIFICITY.  
RC STRAIN=Sprague-Dawley; TISSUE=Adipocyte, and Skeletal muscle;  
RX MEDLINE=94224815; PubMed=81709578;  
RA Hu C., Pang S., Kong X., Velleca M., Lawrence J.C. Jr.;  
RT "Molecular cloning and tissue distribution of PHAS-I, an intracellular  
RT target for insulin and growth factors".  
RL Proc. Natl. Acad. Sci. U.S.A. 91:3730-3734(1994).  
RN [2]  
RP FUNCTION, INTERACTION WITH EIF4E, PHOSPHORYLATION OF SER-64 BY MAPK1  
RP AND MAPK3, AND MUTAGENESIS OF SER-64.  
RX MEDLINE=95025978; PubMed=7939721;  
RA Lin T.-A., Kong X., Haystead T.A.J., Pause A., Belsham G.J.,  
RA Sonenberg N., Lawrence J.C. Jr.;  
RT "PHAS-I as a link between mitogen-activated protein kinase and  
RT translation initiation".  
RL Science 266:653-656(1994).  
CC -1- FUNCTION: Regulates eIF4E activity by preventing its assembly into  
CC the eIF4F complex. Mediates the regulation of protein translation

by hormones, growth factors and other stimuli that signal through the MAP kinase pathway.

-!- SUBUNIT: Nonphosphorylated EIF4EBP1 competes with EIF4G to interact with EIF4E; insulin stimulated MAP-kinase (MAPK1 and MAPK3) phosphorylation of EIF4EBP1 causes dissociation of the complex allowing EIF4G to bind and consequent initiation of translation. Rapamycin can attenuate insulin stimulation, mediated by FKBP.

-!- TISSUE SPECIFICITY: Expressed in all tissues examined; highest levels in fat and skeletal tissue, lowest levels in kidney.

-!- PTM: Phosphorylated on serine and threonine residues in response to insulin, EGF and PDGF.

-!- SIMILARITY: Belongs to the eIF4E binding protein family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U05014; AAA86938.1; -.  
PIR; A55258; A008606; EIF4EBP.  
Pfam; PF05456; EIF4EBP; 1.  
KW Translation regulation; Protein synthesis inhibitor; Phosphorylation.  
FT MOD\_RES 1 1 BLOCKED (PROBABLE).  
FT MOD\_RES 64 64 S->A; DECREASE PHOSPHORYLATION BY MAPK1 AND MAPK3.  
FT MUTAGEN 64 64 R -> N (IN REF. 1; AA SEQUENCE).  
FT CONFLICT 19 19 T -> P (IN REF. 1; AA SEQUENCE).  
FT CONFLICT 69 69 P -> L (IN REF. 1; AA SEQUENCE).  
FT CONFLICT 75 75 P -> L (IN REF. 1; AA SEQUENCE).  
SQ SEQUENCE 117 AA; 12404 MW; 3449D57B09FA101A CRC64;

Query Match 58.1%; Score 25; DB 1; Length 117;  
Best Local Similarity 46.2%; Pred. No. 1.6;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXRXSP 16  
Db 53 YDRKFLMECRNSP 65

RESULT 5  
4BP1\_HUMAN STANDARD; PRT; 118 AA.  
AC Q13541;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Eukaryotic translation initiation factor 4E binding protein 1 (4E-BP1)  
DE (eIF4E-binding protein 1) (phosphorylated heat- and acid-stable protein regulated by insulin 1) (PHAS-I).  
GN EIF4EBP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, INTERACTION WITH EIF4E, AND RP PHOSPHORYLATION.  
RC TISSUE=Placenta;  
RX MEDLINE=95021760; PubMed=7935836;  
RA Pause A., Belsham G.J., Gingras A.-C., Donze O., Lin T.-A., RA Lawrence J.C. Jr., Sonenberg N.;  
RT "Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function.";  
RL Nature 371:762-767(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ito M., Shichijo S., Tsuda N., Ochi M., Harashima N., Saito N.,

RA Itoh K.;  
RT "Identification of multiple genes and immunogenic epitopes of pancreatic cancer cells.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kainane N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S., RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y., RA Phelan M., Farmer A.;  
RT "Cloning of human full-length CDSs in BD Creator (TM) system donor vector.";  
RT Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon, and Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Klausberg R.D., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A., RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C., RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E., RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [5]  
RP INTERACTION WITH EIF4E AND EIF4G;  
RX MEDLINE=96091142; PubMed=8521827;  
RA Haghighat A., Mader S., Pause A., Sonenberg N.;  
RT "Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E.";  
RL EMBO J. 14:5701-5709 (1995).  
CC -!- FUNCTION: Regulates eIF4E activity by preventing its assembly into the EIF4F complex. Mediates the regulation of protein translation by hormones, growth factors and other stimuli that signal through the MAP kinase pathway.  
CC -!- SUBUNIT: Nonphosphorylated EIF4EBP1 competes with EIF4G to interact with EIF4E; insulin stimulated MAP-kinase (MAPK1 and MAPK3) phosphorylation of EIF4EBP1 causes dissociation of the complex allowing EIF4G to bind and consequent initiation of translation. Rapamycin can attenuate insulin stimulation, mediated by FKBP.  
CC -!- PTM: Phosphorylated on serine and threonine residues in response to insulin, EGF and PDGF.  
CC -!- SIMILARITY: Belongs to the eIF4E binding protein family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L36055; AAA62269.1; -.  
EMBL; AB044548; BAB18650.1; -.  
EMBL; BT007162; AAP35826.1; -.  
EMBL; BC004459; AAH04459.1; -.  
EMBL; BC058073; AAH58073.1; -.  
PIR; S50866; S50866.  
DR Genew; HGNC:3288; EIF4EBP1.

DR GK; Q13541; --  
 DR MIM; 602223; --  
 DR GO; GO:0006445; P:regulation of translation; TAS.  
 DR InterPro; IPR008606; EIF4EBP.  
 DR Pfam; PF05456; EIF4EBP; 1.  
 KW Translation regulation; Protein synthesis inhibitor; Phosphorylation.  
 FT MOD\_RES 1 1 BLOCKED (PROBABLE).  
 FT MOD\_RES 65 65 PHOSPHORYLATION (BY MAPK1 AND MAPK3) (BY  
 FT SIMILARITY).

SQ SEQUENCE 118 AA; 12580 MW; 1682AGBA74132966 CRC64;  
 Query Match 58.1%; Score 25; DB 1; Length 118;  
 Best Local Similarity 46.2%; Pred. No. 1.6;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXFLXXXXXSP 16  
 |||||  
 DB 54 YDRKFLMECRNSP 66

RESULT 6  
 ID YMB1 YEAST STANDARD; PRT; 1562 AA.  
 AC Q04781; Q04029;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical 180.2 kDa protein in FAA4-HOR7 intergenic region.  
 GN YMR247C OR YN9408.09C OR YN9920.01C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 ON NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313268; PubMed=9169872;  
 RA Bowman S., Church C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Lyne G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 XII.";  
 RL Nature 387:90-93(1997).  
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
 CC -!- SIMILARITY: STRONG, TO HUMAN ZNF294.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z48639; CAA88574.1; --  
 DR EMBL; Z48756; CAA88657.1; --  
 DR PIR; S53069; S53069.  
 DR Geronline; 142923; --  
 DR SGD; S0004861; YMR247C.  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR001841; Znf ring.  
 DR Pfam; PF00097; ZF-C3HC4; 1.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; FALSE\_NEG.  
 DR PROSITE; PS00889; ZF\_RING\_2; 1.  
 KW Hypothetical protein; Zinc-finger.  
 FT ZN\_FING 1508 1555 RING-TYPE.  
 FT SEQUENCE 1562 AA; 180185 MW; 97AC65E881362305 CRC64;

Query Match 58.1%; Score 25; DB 1; Length 1562;  
 Best Local Similarity 55.6%; Pred. No. 27;  
 Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXXXXSP 16  
 |||||  
 DB 342 FLVSVRTSP 350

RESULT 7  
 ID CXA9 BOVIN STANDARD; PRT; 321 AA.  
 AC Q866T7;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Gap junction alpha-9 protein (Connexin 36) (CX36).  
 GN GJA9.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 ON NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RA Singh V., Pulkuri S., Duda T., Venkataraman V., Mitton K.P.,  
 RA Sitaramayya A.;  
 RT "Bovine connexin-36 (CX36) mRNA";  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: One gap junction consists of a cluster of closely packed  
 pairs of transmembrane channels, the connexons, through which  
 materials of low MW diffuse from one cell to a neighboring cell  
 (By similarity).  
 CC -!- SUBUNIT: A connexon is composed of a hexamer of connexins (By  
 similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the connexin family. Alpha-type (group II)  
 subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AY150575; AAN37413.1; --  
 DR InterPro; IPR000500; Connexin.  
 DR Pfam; PF00029; connexin; 1.  
 DR PRINTS; PR00206; CONNEXIN.  
 DR SMART; SM00037; CNX; 1.  
 DR PROSITE; PS00407; CONNEXINS\_1; FALSE\_NEG.  
 DR PROSITE; PS00408; CONNEXINS\_2; 1.  
 KW Gap junction; Transmembrane.  
 FT DOMAIN 1 19 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 20 42 POTENTIAL.  
 FT DOMAIN 43 75 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 76 98 POTENTIAL.  
 FT DOMAIN 99 197 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 198 220 POTENTIAL.  
 FT DOMAIN 221 252 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 253 275 POTENTIAL.  
 FT DOMAIN 276 321 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 125 138 POLY-GLY.  
 SQ SEQUENCE 321 AA; 36125 MW; 9DD220A10767AB17 CRC64;

Query Match 55.8%; Score 24; DB 1; Length 321;  
 Best Local Similarity 38.5%; Pred. No. 9.7;  
 Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 YXXFLXXXXXSP 16  
 |||||  
 DB 109 YSTVFLALDRDPP 121

```
RESULT 8
CXA9_HUMAN STANDARD; PRT; 321 AA.
ID C9UK14; O9P2R0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Gap junction alpha-9 protein (Connexin 36) (Cx36).
GN GJA9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99393606; PubMed=10462698;
RA Belluaro N., Trovato-Salinaro A., Mudo G., Hurd Y.L.,
RA Condorelli D.F.;
RT 'Structure, chromosomal localization, and brain expression of human
RT Cx36 gene.';
RL J. Neurosci. Res. 57:740-752(1999).
RN [2]
SEQUENCE OF 59-295 FROM N.A.
RA Kitano T., Kobayakawa H., Saitou N.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: One gap junction consists of a cluster of closely packed
CC pairs of transmembrane channels, the connexons, through which
CC materials of low MW diffuse from one cell to a neighboring cell.
CC -!- SUBUNIT: A connexon is composed of a hexamer of connexins.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Highly expressed in neurons.
CC -!- SIMILARITY: Belongs to the connexin family. Alpha-type (group II)
CC subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF153047; AAD54234.1; -.
DR EMBL; AB037509; BAA90429.1; -.
DR MIM; 607058; -.
DR GO; GO:0005921; C:gap junction; NAS.
DR GO; GO:0005243; P:gap-junction forming channel activity; NAS.
DR GO; GO:0007154; P:cell communication; NAS.
DR InterPro; IPR000500; Connexin.
DR Pfam; PF00029; connexin; 1.
DR PRINTS; PR00206; CONNEXIN.
DR SMART; SM00037; CNX; 1.
DR PROSITE; PS00407; CONNEXINS_1; FALSE_NEG.
DR PROSITE; PS00408; CONNEXINS_2; 1.
KW Gap junction; Transmembrane.
FT DOMAIN 1 19 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 20 42 POTENTIAL.
FT DOMAIN 43 75 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 76 98 POTENTIAL.
FT DOMAIN 99 197 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 198 220 POTENTIAL.
FT DOMAIN 221 252 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 253 275 POTENTIAL.
FT DOMAIN 276 321 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 125 138 POLY-GLY.
FT SEQUENCE 321 AA; 36092 MW; 39980B61532A4E76 CRC64;
Query Match 55.8%; Score 24; DB 1; Length 321;
Best Local Similarity 38.5%; Pred. No. 9.7;
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 4 YXXFLXXRXSP 16
```

```
Db 109 YSTVFLALDRDP 121
RESULT 9
YB87_THEME
ID YB87_THEME STANDARD; PRT; 324 AA.
AC Q9X0R9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical protein TM1187.
GN TM1187.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., A.
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0118 [PF0118] FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE001775; AAD36262.1; -.
DR PIR; A72285; A72285.
DR TIGR; TM1187; -.
DR InterPro; IPR002549; UPF0118.
DR Pfam; PF01594; UPF0118; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 190 210 POTENTIAL.
FT TRANSMEM 227 247 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT SEQUENCE 324 AA; 36042 MW; 13D7CA5EA201C327 CRC64;
Query Match 55.8%; Score 24; DB 1; Length 324;
Best Local Similarity 38.5%; Pred. No. 9.8;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 4 YXXFLXXRXSP 16
Db 13 YLVFLVLAKLSP 25
RESULT 10
DCUP_STRAW
ID DCUP_STRAW STANDARD; PRT; 355 AA.
AC Q82KY4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Uroporphyrinogen decarboxylase (EC 4.1.1.37) (URO-D) (UPD).
```

```

GN HEME OR SAV2228.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteridae; Actinobacteriales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis."
RL Nat. Biotechnol. 21:526-531(2003).
RC -!- CATALYTIC ACTIVITY: Uroporphyrinogen-III = coproporphyrinogen + 4
CC CO(2).
CC -!- PATHWAY: Porphyrin biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the uroporphyrinogen decarboxylase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF005030; BAC69939.1; -.
DR HAMAP; MF 00218; -.
DR InterPro; IPR006361; HemeE.
DR InterPro; IPR001638; SBP_bac_3.
DR InterPro; IPR000257; Uro_decaboxyls.
DR Pfam; PF01208; URO-D; 1.
DR ProDom; PD003225; Uro_decaboxyls; 1.
DR TIGRFAMs; TIGR01464; HemeE; 1.
DR PROSITE; PS00906; UROD_1; 1.
DR PROSITE; PS00907; UROD_2; 1.
DR Lyase; Decarboxylase; Porphyrin biosynthesis; Complete proteome.
KW SEQUENCE 355 AA; 38384 MW; 5EC4C01923A4800A CRC64;
SQ SEQUENCE 355 AA; 38384 MW; 5EC4C01923A4800A CRC64;

Query Match 55.8%; Score 24; DB 1; Length 355;
Best Local Similarity 38.5%; Pred. No. 11;
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 YXXFXLXXRXSP 16
DB 16 YDSAFKACRRREP 28

RESULT 11
DCUP STRO
ID DCUP STRO STANDARD; PRT; 355 AA.
AC O69861;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Uroporphyrinogen decarboxylase (EC 4.1.1.37) (URO-D) (UPD).
GN HEME OR SC06031 OR SC1C3.19.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteridae; Actinobacteriales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.

Query Match 55.8%; Score 24; DB 1; Length 355;
Best Local Similarity 38.5%; Pred. No. 11;
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 YXXFXLXXRXSP 16
DB 16 YDSAFKACRRREP 28

RESULT 12
UDB6 RAT
ID UDB6 RAT STANDARD; PRT; 530 AA.
AC P19488;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucuronosyltransferase 2B6 precursor, microsomal (EC 2.4.1.17)
DE (UDPGT) (17-beta-hydroxysteroid specific) (UDPGTR-5).
GN UGT2B6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90256795; PubMed=1692835;
RA McKenzie P.I.;
RT "The cDNA sequence and expression of a variant 17 beta-hydroxysteroid

```

```

RT UDP-glucuronosyltransferase.";
RL J. Biol. Chem. 265:8699-8703(1990).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENDogenous COMPOUNDS. 2B6 IS ABOUT 30-FOLD LESS ACTIVE THAN 2B3
CC TOWARD TESTOSTERONE AND DIHYDROTESTOSTERONE.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
CC beta-D-glucuronoside.
CC -1- INDUCTION: Constitutively expressed.
CC -1- SIMILARITY: Belongs to the UDP-glycosyltransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33746; AAA03216.1; -.
DR EMBL; M33746; AAA03217.1; -.
DR PIR; A36276; A36276.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW Multigene family; Microsome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 530 UDP-GLUCURONOSYLTRANSFERASE 2B6.
FT TRANSMEM 494 510 POTENTIAL.
FT SEQUENCE 530 AA; 60593 MW; F6B23E436B9BDAAE CRC64;
SQ
Query Match 55.8%; Score 24; DB 1; Length 530;
Best Local Similarity 55.8%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 8 FLXXXRXSP 16
DB 169 FLYSLRASP 177
RESULT 13
AGLA RHIME
ID AGLA RHIME STANDARD; PRT; 551 AA.
AC Q9Z3R8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable alpha-glucosidase (EC 3.2.1.20).
GN AGLA OR R00698 OR SMC03064.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99328961; PubMed=10400573;
RA Willis L.B., Walker G.C.;
RT "A novel Sinorhizobium meliloti operon encodes an alpha-glucosidase
RT and a periplasmic-binding-protein-dependent transport system for
RT alpha-glucosides."
RL J. Bacteriol. 181:4176-4184(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=1021.
RC
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont

```

```

RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -1- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF045609; AAD12047.1; -.
DR EMBL; AL591784; CAC45270.1; -.
DR HSSP; P21332; IUOK.
DR InterPro; IPR006589; Alp_amy1_cat_sub.
DR InterPro; IPR006047; Alpha_amy1_cat.
DR Pfam; PF00128; alpha-amy1ase; 1.
DR SMART; SM00642; Aamy; 1.
KW Hydrolase; Glycosidase; Complete proteome.
FT ACT_SITE 212 212 BY SIMILARITY.
FT ACT_SITE 281 281 BY SIMILARITY.
FT ACT_SITE 345 345 BY SIMILARITY.
FT CONFLICT 13 13 P -> A (IN REF. 1).
FT CONFLICT 20 21 GA -> RP (IN REF. 1).
FT CONFLICT 402 415 YGIQWPDFKGRDG -> MASSSGPTSSAGR (IN REF.
FT 1).
FT CONFLICT 445 460 PRAVAVQEGDPASVLH -> RGRCRAGGRPLGAA (IN
FT REF. 1).
SQ SEQUENCE 551 AA; 62576 MW; BB7BD3E17C935509 CRC64;
Query Match 55.8%; Score 24; DB 1; Length 551;
Best Local Similarity 38.5%; Pred. No. 18;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 4 YXXXFLXXXRXSP 16
DB 306 YAFEFAPDRLTP 318
RESULT 14
DCC HUMAN
ID DCC HUMAN STANDARD; PRT; 1447 AA.
AC P43146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).
GN DCC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95011532; PubMed=7926722;
RA Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA Vogelstein B.;
RT "The DCC gene product in cellular differentiation and colorectal
RT tumorigenesis."
RL Genes Dev. 8:1174-1183(1994).
RN [2]
RP SEQUENCE OF 1-750 FROM N.A.
RX MEDLINE=90100559; PubMed=2294591;
RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA Vogelstein B.;
RT "Identification of a chromosome 18q gene that is altered in
RT colorectal cancers."
RL Science 247:49-56(1990).
RN [3]

```



RP SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).  
RX MEDLINE=91121517; PubMed=1991322;  
RA Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Rupbert J.M.,  
RT Oliner J.D., Kinzler K.W., Vogelstein B.;  
RL "Scrambled exons";  
RN Cell 64:607-613(1991).  
[4]  
RP GENE STRUCTURE, AND VARIANT CARCINOMA HIS-1375.  
RX MEDLINE=94245241; PubMed=8188295;  
RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,  
RX Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;  
RT "The DCC gene: structural analysis and mutations in colorectal  
carcinomas";  
RL Genomics 19:525-531(1994).  
[5]  
RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.  
RX MEDLINE=94243823; PubMed=8187090;  
RA Miyake S., Nagai K., Yoshino K., Ota M., Endo M., Yuasa Y.;  
RT "Point mutations and allelic deletion of tumor suppressor gene DCC in  
human esophageal squamous cell carcinomas and their relation to  
metastasis";  
RL Cancer Res. 54:3007-3010(1994).  
CC -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL  
NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.  
CC -!- DISEASE: Colorectal tumors that lost their capacity to  
differentiate into mucus producing cells uniformly lack DCC  
expression. Inactivation of DCC due to allelic deletion and/or  
point mutations may cause both lymphatic and hematogenous  
metastasis of esophageal squamous cell carcinomas.  
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.  
CC -!- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.  
CC -!- SIMILARITY: Contains 6 fibronectin type III domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X76132; CRA53735.1; -  
EMBL; M32292; AAA35751.1; -  
EMBL; M32286; AAA52174.1; -  
EMBL; M32288; AAA52175.1; ALT\_SEQ.  
EMBL; M32290; AAA52176.1; -  
EMBL; M63696; AAA52177.1; -  
EMBL; M63700; AAA52178.1; -  
EMBL; M63702; AAA52179.1; -  
EMBL; M63718; AAA52180.1; -  
EMBL; M63698; AAA52181.1; -  
PIR; A54100; A54100.  
HSSP; P56276; ITLK.  
DCC  
Genew; HGNC:2701; DCC.  
MIM; 120470; -  
GO; GO:0004888; P:transmembrane receptor activity; TAS.  
GO; GO:0007409; P:axogenesis; TAS.  
GO; GO:0006917; P:induction of apoptosis; TAS.  
InterPro; IPR008957; FN III-like.  
InterPro; IPR003961; FN III.  
InterPro; IPR003962; FNIII subd.  
InterPro; IPR007110; Ig-like.  
InterPro; IPR003598; Ig\_c2.  
Pfam; PF00041; fn3; 6.  
Pfam; PF00047; Ig; 4.  
PRINTS; PR00014; FNTYPEIII.  
SMART; SM00060; FN3; 6.  
SMART; SM00408; IGC2; 3.  
PROSITE; PS00835; IG\_LIKE; 4.  
Kw Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;  
Repeat; Anti-oncogene; Disease mutation; Polymorphism.

FT SIGNAL 1 25  
FT CHAIN 26 1447  
FT DOMAIN 26 1097  
FT TRANSMEM 1098 1122  
FT DOMAIN 1123 1447  
FT DOMAIN 26 135  
FT DOMAIN 139 229  
FT DOMAIN 234 326  
FT DOMAIN 331 416  
FT DOMAIN 426 522  
FT DOMAIN 525 618  
FT DOMAIN 619 716  
FT DOMAIN 722 816  
FT DOMAIN 840 940  
FT DOMAIN 941 1042  
FT DISULFID 61 117  
FT DISULFID 161 212  
FT DISULFID 261 310  
FT DISULFID 352 400  
FT CARBOHYD 94 94  
FT CARBOHYD 299 299  
FT CARBOHYD 318 318  
FT CARBOHYD 478 478  
FT CARBOHYD 528 528  
FT CARBOHYD 702 702  
FT VARIANT 168 168  
FT VARIANT 201 201  
FT VARIANT 1375 1375  
FT VARIANT 138 138  
FT CONFLICT 233 329  
FT CONFLICT 421 421  
SQ SEQUENCE 1447 AA; 158456 MW; 4A8612766ED0471F CRC64;  
Query March 55.8%; Score 24; DB 1; Length 1447;  
Best Local Similarity 38.5%; Pred. No. 51;  
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 4 YXXFLXXRXSP 16  
Db 594 YSLRLAYRYGP 606  
RESULT 15  
DCC\_MOUSE  
ID\_DCC\_MOUSE STANDARD; PRT; 1447 AA.  
AC P70211;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Tumor suppressor protein DCC precursor.  
GN DCC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).  
RC STRAIN=BALB/c; TISSUE=Brain;  
RX MEDLINE=96112625; PubMed=8570174;  
RA Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F.;  
RT "Cloning of the mouse homologue of the deleted in colorectal cancer  
gene (mDCC) and its expression in the developing mouse embryo";  
RL Oncogene 11:2243-2254(1995).  
RN [2]  
RP REVISIONS.  
RC STRAIN=BALB/c; TISSUE=Brain;  
RA Cooper H.M.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=A;  
CC IsoId=P70211-1; Sequences=Displayed;  
CC Note=Isoform B is produced by alternative initiation at Met-85  
CC of isoform A;  
CC Name=C;  
CC IsoId=P70211-2; Sequences=VSP\_002501;  
CC Event=Alternative initiation;  
CC Comment=2 isoforms, A (shown here) and B, are produced by  
CC alternative initiation at Met-1 and Met-85;  
CC -!- TISSUE SPECIFICITY: in the embryo, expressed at high levels in the  
CC developing brain and neural tube. In adult, highly expressed in  
CC brain with very low levels found in testis, heart and thymus.  
CC Isoform C is expressed only in the embryo.  
CC -!- DEVELOPMENTAL STAGE: Low levels in early gestation. Highest levels  
CC expressed during mid gestation. Levels decrease in late gestation  
CC and remain at this level in the adult.  
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.  
CC -!- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.  
CC -!- SIMILARITY: Contains 6 fibronectin type III domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X85788; CAA59786.1; -.  
CC HSP; P56276; ITLK.  
CC MGD; MGI:94869; Dcc.  
CC InterPro: IPR008957; FN III-like.  
CC InterPro: IPR003961; FN\_III.  
CC InterPro: IPR003982; FnIII\_subd.  
CC InterPro: IPR007110; Ig-like.  
CC InterPro: IPR003598; Ig\_c2.  
CC Pfam; PF00041; fn3; 6.  
CC Pfam; PF00047; ig; 4.  
CC PRINTS; PR00014; ENTPEIII.  
CC SMART; SM00060; FN3; 6.  
CC SMART; SM00408; IGC2; 3.  
CC PROSITE; PS50835; IG\_LIKE; 4.  
CC Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;  
CC Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM A.  
FT CHAIN 85 1447 TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM B.  
FT INIT MET 85 85 FOR ISOFORM B.  
FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1098 1122 POTENTIAL.  
FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 36 135 IG-LIKE C2-TYPE 1.  
FT DOMAIN 139 229 IG-LIKE C2-TYPE 2.  
FT DOMAIN 234 326 IG-LIKE C2-TYPE 3.  
FT DOMAIN 331 416 IG-LIKE C2-TYPE 4.  
FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.  
FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.  
FT DOMAIN 941 1042 FIBRONECTIN TYPE-III 6.  
FT DISULFID 61 117 BY SIMILARITY.  
FT DISULFID 161 212 BY SIMILARITY.  
FT DISULFID 261 310 BY SIMILARITY.  
FT DISULFID 352 400 BY SIMILARITY.  
FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 819 838 Missing (in isoform C).  
FT /FTId=VSP\_002501.  
SQ SEQUENCE 1447 AA; 158298 MW; OD1FI097C22D5B9F CRC64;  
Query Match 55.8%; Score 24; DB 1; Length 1447;  
Best Local Similarity 38.5%; Pred. No. 51;  
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 4 YXXXFLLXXXKSP 16  
Db 594 YTLRFLAYNRYGP 606  
Search completed: October 5, 2004, 16:07:33  
Job time : 14.1781 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:56:41 ; Search time 62.2466 Seconds  
(without alignments)  
81.102 Million cell updates/sec

Title: US-09-973-473a-24

Perfect score: 43  
Sequence: 1 XXXYXXFLXXXRSP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	69.8	766	12 Q91FV0	Q91fv0 tt virus. p
2	29	67.4	719	12 Q914N0	Q914n0 tt virus. o
3	29	67.4	742	12 Q99AQ9	Q99aq9 tt virus. o
4	28	65.1	727	12 Q9DT02	Q9dtd2 tt virus. o
5	28	65.1	745	12 Q91PS7	Q91ps7 tt virus. o
6	28	65.1	772	12 Q99AR3	Q99ar3 tt virus. a
7	27	62.8	436	16 Q84615	Q84615 chlamydia t
8	27	62.8	721	16 Q8CBP5	Q8cbp5 mycobacteri
9	26	60.5	162	5 Q81LB7	Q81lb7 plasmodium
10	26	60.5	735	12 Q9DUC9	Q9duc9 tt virus. o
11	26	60.5	738	12 Q8V7H5	Q8v7h5 tt virus. o
12	26	60.5	923	5 Q8IQ88	Q8iq88 drosophila
13	26	60.5	1105	10 Q8RVF6	Q8rvf6 oryza sativ
14	25	58.1	102	6 Q8EG57	Q8eg57 sus scrofa
15	25	58.1	109	17 Q50113	Q50113 pyrococcus
16	25	58.1	112	13 Q98TT6	Q98tt6 brachydanio

17	25	58.1	117	5 Q9XZ56	Q9xz56 drosophila
18	25	58.1	143	16 Q9ACU2	Q9acu2 streptomyce
19	25	58.1	163	10 Q84ZC9	Q84zc9 oryza sativ
20	25	58.1	471	10 Q7XUW0	Q7xuw0 oryza sativ
21	25	58.1	547	16 Q92X09	Q92x09 rhizobium m
22	25	58.1	964	4 Q8NHB4	Q8nbb4 homo sapien
23	24	55.8	156	2 Q9F8K0	Q9f8k0 carboxydoch
24	24	55.8	178	6 Q8MKG9	Q8mkg9 saimiri sc1
25	24	55.8	237	6 Q9N2F5	Q9n2f5 pongo pygma
26	24	55.8	237	6 Q9N2F6	Q9n2f6 gorilla gor
27	24	55.8	237	6 Q9N2F7	Q9n2f7 pan troglod
28	24	55.8	244	16 Q88TX5	Q88tx5 lactobacill
29	24	55.8	262	16 Q9CLU0	Q9clu0 pasteurella
30	24	55.8	265	12 Q91BS3	Q91bs3 porcine rep
31	24	55.8	265	12 Q9P283	Q9p283 porcine rep
32	24	55.8	312	16 Q832K3	Q832k3 enterococu
33	24	55.8	329	16 Q7WPO8	Q7wpq8 bordetella
34	24	55.8	329	16 Q7WBR6	Q7wbr6 bordetella
35	24	55.8	350	5 Q8SVL1	Q8svl1 encephalito
36	24	55.8	355	16 Q82KY4	Q82ky4 streptomyce
37	24	55.8	365	8 Q33035	Q33035 resultor m1
38	24	55.8	383	16 Q9RV55	Q9rv55 deinococcus
39	24	55.8	392	4 Q9NYF3	Q9nyf3 homo sapien
40	24	55.8	393	11 Q8BXQ8	Q8bxq8 mus musculu
41	24	55.8	490	13 Q9W6G5	Q9w6g5 brachydanio
42	24	55.8	541	5 Q21484	Q21484 caenorhabdi
43	24	55.8	589	10 Q84QA7	Q84qa7 oryza sativ
44	24	55.8	590	17 Q96XC0	Q96xc0 sulfolobus
45	24	55.8	611	5 Q9V9M6	Q9v9m6 drosophila

ALIGNMENTS

RESULT 1

Q91FV0 ID Q91FV0 PRELIMINARY; PRT; 766 AA.  
AC Q91FV0;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Putative capsid protein.  
OS TT virus.  
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.  
OX NCBI\_TaxID=68887;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PMV;  
RX MEDLINE=20409089; PubMed=10950985;  
RA Hallett R.L., Clewley J.P., Bobet F., McKiernan P.J., Teo C.G.;  
RT "Characterization of a highly divergent TT virus genome.";  
RL J. Gen. Virol. 81:2273-2279(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PMV;  
RX MEDLINE=20409089; PubMed=10950985;  
RA Hallett R.L., Clewley J.P., Bobet F., McKiernan P.J., Teo C.G.;  
RT "Characterization of a highly divergent TT virus genome.";  
RL J. Gen. Virol. 81:2273-2279(2000).

Query Match 69.8%; Score 30; DB 12; Length 766;  
Best Local Similarity 46.2%; Pred. No. 2.1;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXYFLXXXRSP 16

Db 404 YSSIFLANERTSP 416

RESULT 2

Q914N0

```
ID Q914N0 PRELIMINARY; PRT; 719 AA.
AC Q914N0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KAV;
RA Heller F., Zachoval R., Nitschko H., Froesner G.G.;
RT "Isolate KAV: A new member of the TT-virus family.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435014; AAL28134.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 719 AA; 85564 MW; 2AB6912E5B726B38 CRC64;

Query Match 67.4%; Score 29; DB 12; Length 719;
Best Local Similarity 46.2%; Pred. No. 4.1;
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16
Db 381 YSAIFLTNGRLSP 393

RESULT 3
ID Q99AQ9 PRELIMINARY; PRT; 742 AA.
AC Q99AQ9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TCHN-D1;
RA Luo K.-X., He H.-T., Liu D.-X., Liu Z.-H., Xiao H., Jiang X.-J.,
RT Liang W.-F., Zhang L.;
RL "Novel variants related to TT virus wide distribution in China.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF345524; AAK11702.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 742 AA; 87417 MW; 4EE8F24ADAB5BFE9 CRC64;

Query Match 67.4%; Score 29; DB 12; Length 742;
Best Local Similarity 46.2%; Pred. No. 4.3;
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16
Db 402 YSAFLNYHRLSP 414

RESULT 4
ID Q9DTD2 PRELIMINARY; PRT; 727 AA.
AC Q9DTD2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]

ID Q914N0 PRELIMINARY; PRT; 719 AA.
AC Q914N0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=s-TTV CH71;
RA Abe K., Inami T.;
RT "simian TT virus.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB049608; BAB20604.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 727 AA; 85212 MW; 574986EB9239D083 CRC64;

Query Match 65.1%; Score 28; DB 12; Length 727;
Best Local Similarity 38.5%; Pred. No. 8.6;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16
Db 380 YSIFLSSGRSNP 392

RESULT 5
ID Q91PS7 PRELIMINARY; PRT; 745 AA.
AC Q91PS7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kt-08F;
RA Okamoto H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB054647; BAB61607.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 745 AA; 88318 MW; B248C4CFB853900 CRC64;

Query Match 65.1%; Score 28; DB 12; Length 745;
Best Local Similarity 46.2%; Pred. No. 8.8;
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXXRXSP 16
Db 406 YSPIFLSNRLSP 418

RESULT 6
ID Q99AR3 PRELIMINARY; PRT; 772 AA.
AC Q99AR3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ORF1 (Hypothetical protein).
OS TT virus, and
OS TTV-like virus DXL1.
```

OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.  
 RN NCBI\_TaxID=68987, 167759;  
 RP SPECIES=TTV-like virus  
 RC SPECIES=TT virus; STRAIN=TCHN-E;  
 RA Luo K.-X., He H.-T., Liu D.-X., Liu Z.-H., Xiao H., Jiang X.-J.,  
 RA Liang W.-P., Zhang L.;  
 RT "Novel variants related to TTV virus wide distribution in China."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SPECIES=TTV-like virus  
 RC SPECIES=TTV-like virus  
 RA Luo K.-X., He H.-T., Xiao H., Liang W.-P., Liu D.-X.;  
 RT "A novel TTV-like genome detected in both feces and blood from  
 patients in a hepatitis outbreak."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF345522; AAK11698.1; -;  
 DR EMBL; AF315076; AAL37157.1; -;  
 DR InterPro; IPR004219; TTVirus\_Unk.  
 DR Pfam; PF02956; TTV\_ORF1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 772 AA; 90900 MW; 88580F20415FDE3 CRC64;  
 Query Match 65.1%; Score 28; DB 12; Length 772;  
 Best Local Similarity 38.5%; Pred. No. 9.1;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
 QY 4 YXXXFLXXXXXSP 16  
 DB 427 YSSIFLSSGRSNP 439  
 RESULT 7  
 ID 084615 PRELIMINARY; PRT; 436 AA.  
 AC 084615;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE RNA polymerase sigma-54 factor.  
 GN RPON OR C7609.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OC NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UW-3/Cx;  
 RX MEDLINE=99000809; PubMed=9784136;  
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 Chlamydia trachomatis."  
 RL Science 282:754-759(1998).  
 CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES  
 ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND  
 THEN IS RELEASED (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SIGMA-54 FACTOR FAMILY.  
 DR EMBL; AE001331; AAC68212.1; -;  
 DR PIR; E71493; E71493.  
 DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0016987; F:sigma factor activity; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR GO; GO:0006352; P:transcription initiation; IEA.  
 DR InterPro; IPR007046; Sigma54\_CBD.  
 DR InterPro; IPR007634; Sigma54\_DBD.  
 DR Pfam; PF00309; sigma54\_AID; 1.  
 DR Pfam; PF04963; sigma54\_CBD; 1.  
 DR Pfam; PF04552; sigma54\_DBD; 1.  
 DR PRINTS; PR00045; SIGMA54FACT.  
 DR PROSITE; PS00718; SIGMA54\_2; 1.

DR PROSITE; PS50044; SIGMA54\_3; 1.  
 KW DNA-binding; DNA-directed RNA polymerase; Sigma factor;  
 KW Transcription regulation; Complete proteome.  
 SQ SEQUENCE 436 AA; 49150 MW; 7041471872546B4 CRC64;  
 Query Match 62.8%; Score 27; DB 16; Length 436;  
 Best Local Similarity 46.2%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 4 YXXXFLXXXXXSP 16  
 DB 298 YQEFLLKKTSP 310  
 RESULT 8  
 ID 09CBP5 PRELIMINARY; PRT; 721 AA.  
 AC 09CBP5;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein ML1751.  
 GN MLI751.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 403:1007-1011(2001).  
 DR EMBL; AL583923; CAC30704.1; -;  
 DR PIR; A87128; A87128.  
 DR Lepronia; MLI751; -;  
 DR GO; GO:0003824; F:catalytic activity; IEA.  
 DR InterPro; IPR009036; Moeb.  
 DR InterPro; IPR000594; Thif\_domain.  
 DR Pfam; PF00899; Thif; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 721 AA; 78448 MW; 556C2EAA19D65D81 CRC64;  
 Query Match 62.8%; Score 27; DB 16; Length 721;  
 Best Local Similarity 46.2%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 4 YXXXFLXXXXXSP 16  
 DB 281 YILRFLEAERLSP 293  
 RESULT 9  
 ID 08ILB7 PRELIMINARY; PRT; 162 AA.  
 AC 08ILB7;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Mitochondrial import inner membrane translocase subunit tim17,  
 DE putative.  
 GN PF14\_0328.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```

OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=2255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallem S.J., Suh B., Peterson J., Angiuoli S.,
RA Perte M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AF014821; AAN36941.1; -.
DR GO; GO:0005744; C:mitochondrial inner membrane pre-sequence t. . . ; IEA.
DR GO; GO:0015450; F:protein translocase activity; IEA.
DR GO; GO:0015031; P:protein transport; IEA.
DR InterPro; IPR003397; Tim17_Tim22.
DR Pfam; PF02466; Tim17; 1.
SQ SEQUENCE 162 AA; 17909 MW; F536704A462221FF CRC64;

Query Match 60.5%; Score 26; DB 5; Length 162;
Best Local Similarity 46.2%; Pred. No. 8.2;
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXRXSP 16
DB 32 YIWHFLKGARNSP 44

RESULT 10
Q9DUC9 PRELIMINARY; PRT; 735 AA.
AC Q9DUC9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt-TTV6;
RA Okamoto H.; Nishizawa T., Tawara A., Peng Y., Takahashi M.,
RA Kishimoto J., Tanaka T., Miyakawa Y., Mayumi M.;
RT "Species-specific TT viruses in humans and nonhuman primates and their
RT phylogenetic relatedness.";
RL Virology 277:368-378(2000).
DR EMBL; AB041957; BAB19308.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 735 AA; 86132 MW; 9ED18D6BE6FA5D3 CRC64;

Query Match 60.5%; Score 26; DB 12; Length 735;
Best Local Similarity 38.5%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXRXSP 16
DB 392 YSAPFLSAGRLNP 404

```

```

RESULT 11
Q8V7H5 PRELIMINARY; PRT; 738 AA.
AC Q8V7H5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UT05F;
RA Okamoto H.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UT05F;
RX MEDLINE=21844401; PubMed=11855633;
RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,
RA Okamoto H.;
RT "Analysis of the complete genomes of thirteen TT virus variants
RT classifiable into the fourth and fifth genetic groups, isolated from
RT viremic infants.";
RL Arch. Virol. 147:21-41(2002).
DR EMBL; AB064600; BAB79330.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 738 AA; 87631 MW; 57F3032EB1EDE9D2 CRC64;

Query Match 60.5%; Score 26; DB 12; Length 738;
Best Local Similarity 46.2%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 YXXXFLXXRXSP 16
DB 394 YSPFLTPQRYSP 406

RESULT 12
Q8IQ88 PRELIMINARY; PRT; 923 AA.
AC Q8IQ88;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG14837-PB.
GN CG14837.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kimos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celnikier S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikier S.E.,  
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003558; AAN12044.1; -;  
 DR FlyBase; FBgn0035797; CG14837;  
 SQ SEQUENCE 923 AA; 101998 MW; 1380CF09D79CB679 CRC64;  
 Query Match 60.5%; Score 26; DB 5; Length 923;  
 Best Local Similarity 38.5%; Pred. No. 46;  
 Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 Qy 4 YXXFLSSRRSP 876  
 Db 864 YTASLSSRRSP 876

RESULT 13  
 Q8RVF6

ID Q8RVF6 PRELIMINARY; PRT; 1105 AA.  
 AC Q8RVF6;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE B1158C05.16 protein (P0663E10.1 protein).  
 GN B1158C05.16 OR P0663E10.1.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC  
 RT clone:B1158C05.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0663E10.";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003681; BAB90420.1; -;  
 DR EMBL; AP004317; BAB90719.1; -;  
 DR GmGene; Q8RVF6; -;  
 SQ SEQUENCE 1105 AA; 123683 MW; 001F9988334DB9C9 CRC64;  
 Query Match 60.5%; Score 26; DB 10; Length 1105;  
 Best Local Similarity 38.5%; Pred. No. 54;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
 Qy 4 YXXFLXXXRSP 16  
 Db 582 YSLRFISGRDSP 594  
 RESULT 14  
 ID Q8RG57 PRELIMINARY; PRT; 102 AA.  
 AC Q8RG57;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Translation initiation factor 4E binding protein 1 (Fragment).  
 OS Sus scrofa (Fig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Anger M., Klima J., Kubelka M., Carnwath J.W., Niemann H.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF337389; AAK08101.1; -;  
 DR GO; GO:0003743; F:translation initiation factor activity; IEA.  
 DR InterPro; IPR008606; EIF4EBP.  
 DR Pfam; PF05456; EIF4EBP; 1.  
 KW Initiation factor.  
 FT NON\_TER 1  
 FT NON\_TER 102  
 SQ SEQUENCE 102 AA; 10697 MW; 01C88769D02658E9 CRC64;  
 Query Match 58.1%; Score 25; DB 6; Length 102;  
 Best Local Similarity 46.2%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 Qy 4 YXXFLXXXRSP 16  
 Db 44 YDRFLMECRNSP 56

```
RESULT 15
O50113
ID O50113 PRELIMINARY; PRT; 109 AA.
AC O50113;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PH1405.
GN PH1405.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb P.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RL thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000006; BAA30511.1; -.
DR PIR; G71013; G71013.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 109 AA; 11411 MW; CDFOD2FE3A1BD32D CRC64;

Query Match 58.1%; Score 25; DB 17; Length 109;
Best Local Similarity 55.6%; Pred. No. 11;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 FLXXXRXSP 16
Db |||||
17 FLASINTSP 25
```

Search completed: October 5, 2004, 16:12:32  
Job time : 67.2466 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:06 ; Search time 82.1918 Seconds  
(without alignments)  
55.003 Million cell updates/sec

Title: US-09-973-473A-26  
Perfect score: 30  
Sequence: 1 XXXYXXFLXXXXXX 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	17	56.7	18	3	AAY77921	Aay77921 HERG gene
2	17	56.7	41	4	AAM21233	Aam21233 Peptide #
3	17	56.7	41	4	ABB43559	Abb43559 Peptide #
4	17	56.7	41	4	AAM37455	Aam37455 Peptide #
5	17	56.7	41	4	ABB26517	Abb26517 Protein #
6	17	56.7	41	4	AAM77299	Aam77299 Human Don
7	17	56.7	41	4	AAM64495	Aam64495 Human liv
8	17	56.7	41	4	ABG58934	Abg58934 Human bra
9	17	56.7	41	5	ABG46317	Abg46317 Human pep
10	17	56.7	44	4	ABB15751	Abb15751 Human ner
11	17	56.7	57	5	AU75474	Aau75474 Human gon
12	17	56.7	57	7	ADC71253	Adc71253 Human col
13	17	56.7	86	5	ABP08557	Abp08557 Human ORF
14	17	56.7	86	5	AU79208	Aau79208 Human pro
15	17	56.7	105	5	ABP06736	Abp06736 Human ORF
16	17	56.7	107	5	AAE19664	Aae19664 Human TNF
17	17	56.7	107	5	AAE19665	Aae19665 Human TNF
18	17	56.7	107	6	ABR55856	Abri55856 Anti-Her-
19	17	56.7	108	5	AAE19684	Aae19684 Mouse-hum
20	17	56.7	109	2	AAE19684	Aae19684 Mouse-hum
21	17	56.7	109	2	AAE19684	Aae19684 Mouse-hum
22	17	56.7	109	2	AAE19684	Aae19684 Mouse-hum
23	17	56.7	109	2	AAE19684	Aae19684 Mouse-hum
24	17	56.7	109	2	AAE19684	Aae19684 Mouse-hum
25	17	56.7	109	2	AAE19684	Aae19684 Mouse-hum

26	17	56.7	149	5	ABB89878	Abb89878 Human pol
27	17	56.7	154	5	ABU05821	Abu05821 M. tuberc
28	17	56.7	156	5	ABU05414	Abu05414 M. tuberc
29	17	56.7	162	7	ADC94479	Adc94479 E. faeciu
30	17	56.7	166	2	AAR22133	Aar22133 Sequence
31	17	56.7	166	7	ADB65149	Adb65149 Human pro
32	17	56.7	170	3	AAE96467	Aae96467 Partial V
33	17	56.7	178	3	AAE53256	Aae53256 Human col
34	17	56.7	178	4	AAE74057	Aae74057 Human col
35	17	56.7	185	5	AAE21463	Aae21463 Human gen
36	17	56.7	185	5	ABG64962	Abg64962 Human alb
37	17	56.7	187	5	ABP41825	Abp41825 Human ova
38	17	56.7	192	3	AAE41568	Aae41568 Arabidops
39	17	56.7	195	3	AAE41567	Aae41567 Arabidops
40	17	56.7	209	4	AAE85266	Aae85266 Human imm
41	17	56.7	209	4	AAE42383	Aae42383 Human pol
42	17	56.7	212	2	AAW20152	Aaw20152 H. pylori
43	17	56.7	214	5	AAE19696	Aae19696 Antibody
44	17	56.7	214	7	ADB85319	Adb85319 Light cha
45	17	56.7	214	7	ADC26166	Adc26166 Humanised

ALIGNMENTS

RESULT 1  
AAY77921  
ID AAY77921 standard; peptide; 18 AA.  
XX  
AC AAY77921;  
XX  
DT 19-JUN-2000 (first entry)  
XX  
DE HERG gene intragenic fragment.  
XX  
KW HERG; mutation; long QT syndrome; LQT syndrome; gene therapy; deletion;  
KW human.  
XX  
OS Homo sapiens.  
XX  
PN WO200006772-A1.  
XX  
PD 10-FEB-2000.  
XX  
PF 20-JUL-1999; 99WO-US016337.  
XX  
PR 27-JUL-1998; 98US-00122847.  
XX  
PR 06-JAN-1999; 99US-00226012.  
XX  
(UTAH ) UNIV UTAH RES FOUND.  
XX  
PI Keating MT, Splawski I;  
XX  
DR WPI; 2000-195319/17.  
XX  
DR N-PSDB; AAA07695.  
XX  
PT New isolated mutant HERG nucleic acids, useful for developing products  
XX for the diagnosis, prevention and treatment of long QT syndrome.  
XX  
PS Disclosure; Fig 11C; 163pp; English.  
XX  
CC The invention relates to a HERG protein having a mutation compared to  
XX wild-type HERG, and is useful for developing products for the diagnosis,  
XX prevention and treatment of long QT (LQT) syndrome. The products and  
XX methods can be used for the diagnosis of subjects with LQT syndrome. They  
XX can also be used to screen for drugs for treating or preventing LQT  
XX syndrome. The HERG nucleic acids can also be used for gene therapy and  
XX HERG peptides can be used for peptide therapy. The present sequence  
XX represents a HERG fragment  
SQ Sequence 18 AA;

Query Match 56.7%; Score 17; DB 3; Length 18;

```
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 10 YSAAFL 15

RESULT 2
AAM21233
ID AAM21233 standard; protein; 41 AA.
XX AC
XX AAM21233;
XX 12-OCT-2001 (first entry)
XX DE Peptide #7667 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX OS Homo sapiens.
XX XX WO200157278-A2.
XX PN 09-AUG-2001.
XX PD
XX PF 30-JAN-2001; 2001WO-US000670.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488901/53.
XX DR WPI; 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human cervical epithelial cells.
XX PS Claim 27; SEQ ID NO 26059; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX CC can be used to produce a single exon microarray, which can be used for
XX CC measuring human gene expression in a sample derived from human cervical
XX CC epithelial cells. By measuring gene expression, the probes are therefore
XX CC useful in grading and/or staging of diseases of the cervix, notably
XX CC cervical cancer. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 41 AA;

Query Match 56.7%; Score 17; DB 4; Length 41;
Best Local Similarity 50.0%; Pred. No. 5.7e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 21 YSTSFL 26

RESULT 3
ABB43559
ID ABB43559 standard; peptide; 41 AA.
XX AC
XX AAM37455
XX 17-OCT-2001 (first entry)
XX DE Peptide #11492 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX XX WO200157272-A2.
XX PN
```

```
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 37724; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs;
XX see A131315-A157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
XX SQ Sequence 41 AA;
XX
XX Query Match 56.7%; Score 17; DB 4; Length 41;
XX Best Local Similarity 50.0%; Pred. No. 5.7e+02;
XX Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 4 YXXXPFL 9
XX | | |
XX 21 YSTSPFL 26
XX
XX
XX RESULT 5
XX ABB26517
XX ID ABB26517 standard; protein; 41 AA.
XX
XX AC ABB26517;
XX
XX DT 23-JAN-2002 (first entry)
XX
XX DE Protein #8516 encoded by probe for measuring heart cell gene expression.
XX
XX KW Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX
XX OS Homo sapiens.
XX
XX PN WO200157274-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000666.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX
XX PR 26-MAY-2000; 2000US-0207456P.
XX
XX PR 30-JUN-2000; 2000US-00608408.
XX
XX PR 03-AUG-2000; 2000US-00632366.
XX
XX PR 21-SEP-2000; 2000US-0234687P.
XX
XX PR 27-SEP-2000; 2000US-0236359P.
XX
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX
```

```
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX PS Claim 15; SEQ ID NO 28287; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 41 AA;
XX
XX Query Match 56.7%; Score 17; DB 4; Length 41;
XX Best Local Similarity 50.0%; Pred. No. 5.7e+02;
XX Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 4 YXXXPFL 9
XX | | |
XX 21 YSTSPFL 26
XX
XX
XX RESULT 6
XX AAM77299
XX ID AAM77299 standard; protein; 41 AA.
XX
XX AC AAM77299;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 37605.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000668.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX
XX PR 26-MAY-2000; 2000US-0207456P.
XX
XX PR 30-JUN-2000; 2000US-00608408.
XX
XX PR 03-AUG-2000; 2000US-00632366.
XX
XX PR 21-SEP-2000; 2000US-0234687P.
XX
XX PR 27-SEP-2000; 2000US-0236359P.
XX
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
XX PS Example 4; SEQ ID NO 37605; 658pp + Sequence Listing; English.
XX
XX
```

CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention

XX SQ Sequence 41 AA;

Query Match 56.7%; Score 17; DB 4; Length 41;  
 Best Local Similarity 50.0%; Pred. No. 5.7e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
 | | |  
 DB 21 YSTSFL 26

RESULT 7  
 AAM64495  
 ID AAM64495 standard; protein; 41 AA.  
 XX  
 AC AAM64495;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 36600.  
 XX  
 DE Human; brain expressed exon; gene expression analysis; probe; microarray;  
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.

XX  
 PF 30-JAN-2001; 2001WO-US000667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains.

XX Example 4; SEQ ID NO 36600; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention

XX SQ Sequence 41 AA;

Query Match 56.7%; Score 17; DB 4; Length 41;  
 Best Local Similarity 50.0%; Pred. No. 5.7e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9

DB 21 YSTSFL 26

RESULT 8

ABG58934  
 ID ABG58934 standard; peptide; 41 AA.

XX  
 AC ABG58934;

XX DT 25-FEB-2003 (first entry)

XX DE Human liver peptide, SEQ ID No 37582.

XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.

XX OS Homo sapiens.

XX PN WO200157273-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000664.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 WPI; 2001-488898/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.

XX Claim 27; SEQ ID NO 37582; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SEN) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 41 AA;

Query Match 56.7%; Score 17; DB 4; Length 41;  
 Best Local Similarity 50.0%; Pred. No. 5.7e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
 | | |  
 DB 21 YSTSFL 26

RESULT 9

ABG46317  
 ID ABG46317 standard; peptide; 41 AA.

XX ABG46317;  
XX 19-AUG-2002 (first entry)  
XX Human peptide encoded by genome-derived single exon probe SEQ ID 35982.  
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
XX chronic obstructive pulmonary disease; interstitial lung disease;  
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;  
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
XX primary ciliary dyskinesia; pulmonary hypertension;  
XX hyaline membrane disease.  
XX Homo sapiens.  
XX WO200186003-A2.  
XX 15-NOV-2001.  
XX 30-JAN-2001; 2001WO-US000665.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.  
XX Claim 27; SEQ ID NO 35982; 634pp; English.  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human lung comprising single exon nucleic acid probes having one of  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX complements or the 12387 open reading frames derived from the 12614  
XX probes. Also included are a microarray comprising the novel set of probes  
XX; the novel set of probes which hybridize at high stringency to a nucleic  
XX acid expressed in the human lung; measuring gene expression in a sample  
XX derived from human lung, comprising (a) contacting the array with a  
XX collection of detectably labeled nucleic acids derived from human lung  
XX mRNA, and (b) measuring the label detectably bound to each probe of the  
XX array; identifying exons in a eukaryotic genome, comprising (a)  
XX algorithmically predicting at least one exon from genomic sequences of  
XX the eukaryote; and (b) detecting specific hybridization of detectably  
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
XX having a fragment identical to the predicted exon, the probe is included  
XX in the above mentioned microarray; assigning exons to a single gene,  
XX comprising (a) identifying exons from genomic sequence by the method  
XX above and (b) measuring the expression of each of the exons in several  
XX tissues and/or cell types using hybridization to a single exon  
XX microarrays having a probe with the exon, where a common pattern of  
XX expression of the exons in the tissues and/or cell types indicates that  
XX the exons should be assigned to a single gene; a peptide comprising one  
XX of 12011 sequences, mentioned in the specification, or encoded by the  
XX probes/open reading frames (ORF). The probes are used for gene expression  
XX analysis, and for identifying exons in a gene, particularly using human  
XX lung derived mRNA and for the study of lung diseases such as asthma, lung  
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,

CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 41 AA;

Query Match 56.7%; Score 17; DB 5; Length 41;

Best Local Similarity 50.0%; Fred. No. 5.7e+02;

Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXKFL 9

Db 21 YSTSFL 26

RESULT 10

ABBI5751

ID ABBI5751 standard; protein; 44 AA.

XX AC ABBI5751;

XX DT 23-JAN-2002 (first entry)

XX DE Human nervous system related polypeptide SEQ ID NO 4408.

KW Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antisickling; antianemic; antithrictic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

XX OS Homo sapiens.

XX FN WO200159063-A2.

XX PD 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US001334.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.



Matches	3;	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
QY	4	YXXXFL 9							
DB	39	YTTTFL 44							
RESULT 11									
AAU75474									
ID		AAU75474		standard; protein; 57	AA.				
XX		AC							
XX		AAU75474;							
XX		23-APR-2002		(first entry)					
XX									
XX		Human gonadotropin releasing hormone receptor-like GPCR.							
XX		Human; receptor; gonadotropin releasing hormone-like; receptor;							
KW		G protein-coupled receptor; GPCR; GnRH; urinary incontinence;							
KW		benign prostatic hyperplasia; obesity; cancer; diabetes; pain;							
KW		osteoporosis; schizophrenia; neurodegenerative disorder; asthma;							
KW		Parkinson's disease; Alzheimer's disease; acute heart failure;							
KW		angina pectoris; myocardial infarction; dyskinesia; Huntington's disease;							
KW		Tourette's syndrome; infection; human immunodeficiency virus infection.							
XX									
XX		Homo sapiens.							
OS									
XX		WO200200701-A2.							
PN									
XX		03-JAN-2002.							
PD									
XX									
XX		25-JUN-2001; 2001WO-EP007212.							
PF									
XX									
XX		26-JUN-2000; 2000US-0214249P.							
PR									
XX		(FARB ) BAYER AG.							
PA									
XX									
XX		Ramakrishnan S;							
PI									
XX		WPI; 2002-083336/11.							
DR		N-PSDB; ABK13400.							
DR									
XX									
PT		New gonadotropin releasing hormone receptor-like GPCR polypeptide useful							
PT		for the treatment of for example urinary incontinence, obesity and							
PT		related disease and cancer.							
XX									
XX		Claim 25; Fig 2; 96pp; English.							
PS									
CC									
CC		The invention relates to a purified polypeptide comprising a gonadotropin							
CC		releasing hormone receptor (GnRH)-like GPCR (G protein-coupled receptor),							
CC		its encoding polynucleotide, fragments, homologues and allelic variants.							
CC		Also included are a expression vector comprising the polynucleotide, a							
CC		host cell comprising the vector and identification and isolation of							
CC		modulators of the protein and the polynucleotide. The protein,							
CC		polynucleotide and identified modulators are useful in the diagnosis and							
CC		treatment of urinary incontinence, benign prostate hyperplasia, obesity							
CC		and related disease, cancer, diabetes, osteoporosis, anxiety, depression,							
CC		hypertension, migraine, compulsive disorder, schizophrenia, autism,							
CC		neurodegenerative disorder (e.g. Parkinson's disease and Alzheimer's							
CC		disease), acute heart failure, hypotension, angina pectoris, myocardial							
CC		infarction, ulcer, asthma, allergy, delirium, dementia, mental							
CC		retardation, dyskinesia (e.g. Huntington's disease and Tourette's							
CC		syndrome), cachexia, chemotherapy induced vomiting, infections							
CC		(bacterial, viral e.g. human immunodeficiency (HIV) virus, fungal, and							
CC		protozoan), pain, anorexia, bulimia and asthma. The present sequence							
CC		represents the GnRH-like GPCR							
XX									
XX		Sequence 57 AA;							
Query Match		56.7%;		Score 17;	DB 5;	Length 57;			
Best Local Similarity		50.0%;		Pred. No. 7.7e+02;					
Matches		3;		Conservative	0;	Mismatches	3;		

DE Human ORFX protein sequence SEQ ID NO:17096.  
XX Human; open reading frame; ORFX, gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX Homo sapiens.  
OS WO200192523-A2.  
XX  
PN 06-DEC-2001.  
PD  
XX  
XX 29-MAY-2001; 2001WO-US010836.  
PF  
XX 30-MAY-2000; 2000US-0206132P.  
PR 29-AUG-2000; 2000US-0228716P.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX Shimkets RA, Leach MD;  
PI  
XX WPI: 2002-106308/14.  
DR N-PSDB; ABN24309.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.  
XX  
PS Disclosure; SEQ ID NO 17096; 1037pp; English.  
XX  
CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 86 AA;

Query Match 56.7%; Score 17; DB 5; Length 86;  
Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
Db 36 YATTF 41

RESULT 14  
AAU79208

ID AAU79208 standard; protein; 86 AA.  
XX  
AC AAU79208;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human protein 9 similar to inclusin.  
XX  
KW Human; protein 9; inclusin; cytostatic; virucide; haemostatic;  
KW immunomodulator; antiinflammatory; malignant tumour; cancer; haemopathy;  
KW development disorder; HIV infection; immunological disease; inflammation;  
KW growth development disorder; human immunodeficiency virus;  
KW endocrinopathy.  
XX  
OS Homo sapiens.  
XX  
PN WO200183686-A2.  
XX  
PD 08-NOV-2001.  
XX  
XX 28-APR-2001; 2001WO-CN000642.  
PF  
XX 29-APR-2000; 2000CN-00115525.  
PR (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.  
XX  
XX Mao Y, Xie Y;  
PI  
XX WPI: 2002-062118/08.  
DR N-PSDB; ABK49279.  
XX  
XX Human protein 9 similar to inclusin and encoded polynucleotide,  
PT applicable in diagnosis and treatment of developmental disorders.  
PT Malignant tumor, hemopathy, HIV infection, immunological diseases and  
PT various inflammations.  
XX  
PS Claim 1; Page 31; 33pp; Chinese.  
XX  
CC The invention relates to the human protein 9 similar to inclusin and the  
CC polynucleotide encoding it. The polypeptide and encoded polynucleotide  
CC are applicable in diagnosis and treatment of malignant tumours, and  
CC development disorders, haemopathy, HIV infection, immunological diseases,  
CC various inflammations, growth development disorders and endocrinopathy.  
CC This sequence represents human protein 9 similar to inclusin  
XX  
SQ Sequence 86 AA;  
Query Match 56.7%; Score 17; DB 5; Length 86;  
Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4 YXXXFL 9  
Db 33 YSASF 38  
RESULT 15  
ABP06736  
ID ABP06736 standard; protein; 105 AA.  
XX  
AC ABP06736;  
XX  
DT 24-JUN-2002 (first entry)  
XX  
DE Human ORFX protein sequence SEQ ID NO:13454.  
XX  
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;



KW myasthenia gravis.  
XX  
OS Homo sapiens.  
XX  
PN WO200192523-A2.  
XX  
PD 06-DEC-2001.  
XX  
PF 29-MAY-2001; 2001WO-US010836.  
XX  
PR 30-MAY-2000; 2000US-0206132P.  
PR 29-AUG-2000; 2000US-0228716P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach MD;  
XX  
DR WPI; 2002-106308/14.  
DR N-PSDB; AEN22488.  
XX  
PT Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.  
XX  
PS Disclosure; SEQ ID NO 13454; 1037pp; English.  
XX  
CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). AEN15762 to AEN27252 encode the human ORFX  
CC proteins given in AEP00010 to AEP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis.  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 105 AA;  
  
Query Match 56.7%; Score 17; DB 5; Length 105;  
Best Local Similarity 50.0%; Pred. No. 1.4e+03;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 YXXFL 9  
Db 62 YTSFL 67  
  
Search completed: October 5, 2004, 16:06:44  
Job time : 87.1918 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:00:17 ; Search time 21.6986 Seconds  
(without alignments)  
38.068 Million cell updates/sec

Title: US-09-973-473A-26  
Perfect score: 30  
Sequence: 1 XXXXXXXFLXXXXXX 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/ECTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	56.7	18	1	US-08-401-512-64
2	17	56.7	18	3	US-09-226-012-99
3	17	56.7	107	2	US-07-934-373C-41
4	17	56.7	107	2	US-07-934-373C-43
5	17	56.7	107	3	US-08-437-642B-41
6	17	56.7	107	3	US-08-437-642B-43
7	17	56.7	109	2	US-07-934-373C-1
8	17	56.7	109	3	US-08-437-642B-1
9	17	56.7	109	4	US-08-146-206C-1
10	17	56.7	109	4	US-09-705-686-1
11	17	56.7	109	5	PCT-US93-07832-1
12	17	56.7	162	4	US-09-107-532A-4106
13	17	56.7	166	3	US-08-765-381-12
14	17	56.7	214	4	US-09-678-397-1
15	17	56.7	214	4	US-09-680-148-1
16	17	56.7	214	4	US-09-304-465A-1
17	17	56.7	222	4	US-09-489-039A-9160
18	17	56.7	225	2	US-09-489-039A-9780
19	17	56.7	237	2	US-08-463-587A-25
20	17	56.7	237	2	US-08-463-667A-3
21	17	56.7	237	3	US-08-923-854-25
22	17	56.7	237	5	PCT-US91-09133-26
23	17	56.7	310	4	US-09-198-452A-864
24	17	56.7	310	4	US-09-543-681A-6858
25	17	56.7	357	4	US-09-404-286B-10
26	17	56.7	392	4	US-09-107-532A-5078
27	17	56.7	412	4	US-09-489-039A-13259

28	17	56.7	453	4	US-09-107-532A-4019	Sequence 4019, Ap
29	17	56.7	573	4	US-09-328-352-4675	Sequence 4675, Ap
30	17	56.7	594	4	US-09-252-991A-27749	Sequence 27749, A
31	17	56.7	604	4	US-09-820-809-13	Sequence 13, Appl
32	17	56.7	626	2	US-08-956-242-2	Sequence 2, Appl
33	17	56.7	626	3	US-09-351-215-2	Sequence 2, Appl
34	17	56.7	664	3	US-09-268-140-2	Sequence 2, Appl
35	17	56.7	694	4	US-09-489-039A-9066	Sequence 9066, Ap
36	17	56.7	708	4	US-09-489-039A-11984	Sequence 11984, A
37	17	56.7	888	2	US-08-956-242-4	Sequence 4, Appl
38	17	56.7	888	3	US-09-351-215-4	Sequence 4, Appl
39	17	56.7	912	3	US-08-617-785-2	Sequence 2, Appl
40	17	56.7	912	4	US-09-641-318-2	Sequence 2, Appl
41	17	56.7	912	4	US-09-817-464-2	Sequence 2, Appl
42	17	56.7	912	5	PCT-US91-09422-19	Sequence 19, Appl
43	17	56.7	1159	2	US-08-956-242-13	Sequence 13, Appl
44	17	56.7	1159	3	US-09-351-215-13	Sequence 13, Appl
45	17	56.7	1159	3	US-09-226-012-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-401-512-64  
; Sequence 64, Application US/08401512  
; Patent No. 5599673  
; GENERAL INFORMATION:  
; APPLICANT: Keating, Mark T.  
; APPLICANT: Curran, Mark E.  
; TITLE OF INVENTION: Long QT Syndrome Genes  
; NUMBER OF SEQUENCES: 81  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
; STREET: 1201 New York Avenue, Suite 1000  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3917  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/401,512  
; FILING DATE: 09-MAR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saxe, Stephen A.  
; REGISTRATION NUMBER: 38,609  
; REFERENCE/DOCKET NUMBER: 19780-113879  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4848  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
US-08-401-512-64

Query Match 56.7%; Score 17; DB 1; Length 18;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFPL 9  
|||

```
Db      10 YSAAFL 15

RESULT 2
US-09-226-012-99
; Sequence 99, Application US/09226012
; Patent No. 6207383
; GENERAL INFORMATION:
; APPLICANT: Keating, Mark T.
; APPLICANT: Splawski, Igor
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT
; FILE REFERENCE: 2323-136
; CURRENT APPLICATION NUMBER: US/09/226,012
; EARLIER APPLICATION NUMBER: 09/122,847
; EARLIER FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 99
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-226-012-99

Query Match      56.7%; Score 17; DB 3; Length 18;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      | |||
Db      10 YSAAFL 15

RESULT 3
US-07-934-373C-41
; Sequence 41, Application US/07934373C
; Patent No. 5821337
; GENERAL INFORMATION:
; APPLICANT: Paul J. Carter
; APPLICANT: Leonard G. Presta
; TITLE OF INVENTION: Immunoglobulin Variants
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/934,373C
; FILING DATE: 21-Aug-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/05126
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/715272
; FILING DATE: 14-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P0709P2
; TELEPHONE: 650/225-1994
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 107 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-07-934-373C-41

Query Match      56.7%; Score 17; DB 2; Length 107;
Best Local Similarity 50.0%; Pred. No. 5.8e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      | |||
Db      49 YSASF 54

RESULT 5
US-08-437-642B-41
; Sequence 41, Application US/08437642B
; Patent No. 6054297
```

; GENERAL INFORMATION:  
 ; APPLICANT: Paul J. Carter  
 ; APPLICANT: Leonard G. Presta  
 ; TITLE OF INVENTION: Immunoglobulin Variants  
 ; NUMBER OF SEQUENCES: 47  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 1 DNA Way  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WinPatIn (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/437,642B  
 ; FILING DATE: 09-May-1995  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/934373  
 ; FILING DATE: 21-AUG-1992  
 ; APPLICATION NUMBER: 08/146206  
 ; FILING DATE: 17-NOV-1993  
 ; APPLICATION NUMBER: PCT/US92/05126  
 ; FILING DATE: 15-JUN-1992  
 ; APPLICATION NUMBER: PCT/US92/05126  
 ; FILING DATE: 15-JUN-1992  
 ; APPLICATION NUMBER: 07/715272  
 ; FILING DATE: 14-JUN-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER: 40,378  
 ; REFERENCE/DOCKET NUMBER: P0709P2C1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650/225-1994  
 ; TELEFAX: 650/952-9881  
 ; INFORMATION FOR SEQ ID NO: 41:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 107 amino acids  
 ; TYPE: Amino Acid  
 ; TOPOLOGY: Linear  
 ; US-08-437-642B-41

Query Match 56.7%; Score 17; DB 3; Length 107;  
 Best Local Similarity 50.0%; Pred. No. 5.8e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9  
 DB 49 YSASFL 54

RESULT 6  
 ; US-08-437-642B-43  
 ; Sequence 43, Application US/08437642B  
 ; Patent No. 6054297  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Paul J. Carter  
 ; APPLICANT: Leonard G. Presta  
 ; TITLE OF INVENTION: Immunoglobulin Variants  
 ; NUMBER OF SEQUENCES: 47  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 1 DNA Way  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WinPatIn (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/437,642B  
 ; FILING DATE: 09-May-1995  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/934373  
 ; FILING DATE: 21-AUG-1992  
 ; APPLICATION NUMBER: 08/146206  
 ; FILING DATE: 17-NOV-1993  
 ; APPLICATION NUMBER: PCT/US92/05126  
 ; FILING DATE: 15-JUN-1992  
 ; APPLICATION NUMBER: PCT/US92/05126  
 ; FILING DATE: 15-JUN-1992  
 ; APPLICATION NUMBER: 07/715272  
 ; FILING DATE: 14-JUN-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER: 40,378  
 ; REFERENCE/DOCKET NUMBER: P0709P2C1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650/225-1994  
 ; TELEFAX: 650/952-9881  
 ; INFORMATION FOR SEQ ID NO: 43:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 107 amino acids  
 ; TYPE: Amino Acid  
 ; TOPOLOGY: Linear  
 ; US-08-437-642B-43

Query Match 56.7%; Score 17; DB 3; Length 107;  
 Best Local Similarity 50.0%; Pred. No. 5.8e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9  
 DB 49 YSASFL 54

RESULT 7  
 ; US-07-934-373C-1  
 ; Sequence 1, Application US/07934373C  
 ; Patent No. 5821337  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Paul J. Carter  
 ; APPLICANT: Leonard G. Presta  
 ; TITLE OF INVENTION: Immunoglobulin Variants  
 ; NUMBER OF SEQUENCES: 48  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 1 DNA Way  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WinPatIn (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/934,373C  
 ; FILING DATE: 21-Aug-1992  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/05126  
 ; FILING DATE: 15-JUN-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/715272

```
/ FILING DATE: 14-JUN-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Lee, Wendy M.
/ REGISTRATION NUMBER: 40,378
/ REFERENCE/DOCKET NUMBER: P0709P2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650/225-1994
/ TELEFAX: 650/952-9881
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 109 amino acids
/ TYPE: Amino Acid
/ TOPOLOGY: Linear
/ US-07-934-373C-1

Query Match 56.7%; Score 17; DB 2; Length 109;
Best Local Similarity 50.0%; Pred. No. 5.9e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 49 YSASFL 54

RESULT 8
US-08-437-642B-1
; Sequence 1, Application US/08437642B
; Patent No. 6054297
; GENERAL INFORMATION:
; APPLICANT: Paul J. Carter
; APPLICANT: Leonard G. Presta
; TITLE OF INVENTION: Immunoglobulin Variants
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437,642B
; FILING DATE: 09-May-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934373
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION NUMBER: 08/146206
; FILING DATE: 17-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/05126
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/715272
; FILING DATE: 14-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P0709P2C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1994
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear

Query Match 56.7%; Score 17; DB 2; Length 109;
Best Local Similarity 50.0%; Pred. No. 5.9e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 49 YSASFL 54

RESULT 9
US-08-146-206C-1
; Sequence 1, Application US/08146206C
; Patent No. 6407213
; GENERAL INFORMATION:
; APPLICANT: Carter, Paul J.
; APPLICANT: Presta, Leonard G.
; TITLE OF INVENTION: Method for Making Humanized Antibodies
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/146,206C
; FILING DATE: 17-No. 6407213-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/715272
; FILING DATE: 14-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P0709P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1994
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-146-206C-1

Query Match 56.7%; Score 17; DB 4; Length 109;
Best Local Similarity 50.0%; Pred. No. 5.9e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 49 YSASFL 54

RESULT 10
US-09-705-686-1
; Sequence 1, Application US/09705686
; Patent No. 6639055
; GENERAL INFORMATION:
; APPLICANT: Carter, Paul J.
; APPLICANT: Presta, Leonard G.
; TITLE OF INVENTION: Method for Making Humanized Antibodies
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
```

STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/705,686  
FILING DATE: 02-Nov-1993  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/146206  
FILING DATE: 17-NOV-1993  
APPLICATION NUMBER: 07/715272  
FILING DATE: 14-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 40,378  
REFERENCE/DOCKET NUMBER: P0709PID3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1994  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 109 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-705-686-1

Query Match 56.7%; Score 17; DB 4; Length 109;  
Best Local Similarity 50.0%; Pred. No. 5.9e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9  
DB 49 YSASF 54

RESULT 11  
PCT-US93-07832-1  
Sequence 1, Application PC/TUS9307832  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
TITLE OF INVENTION: Immunoglobulin Variants  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07832  
FILING DATE: 19930820  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/715272  
FILING DATE: 14-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/05126  
FILING DATE: 15-JUN-1992  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/934373  
FILING DATE: 21-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: 709P2PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE:  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 109 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US93-07832-1

Query Match 56.7%; Score 17; DB 5; Length 109;  
Best Local Similarity 50.0%; Pred. No. 5.9e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9  
DB 49 YSASF 54

RESULT 12  
US-09-107-532A-4106  
Sequence 4106, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSER: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 162 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature

```
/ LOCATION: (B) LOCATION 1...162
/ SEQUENCE DESCRIPTION: SEQ ID NO: 4106:
US-09-107-532A-4106

Query Match          56.7%; Score 17; DB 4; Length 162;
Best Local Similarity 50.0%; Pred. No. 8.3e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 43 YTA AFL 48

RESULT 13
US-08-765-381-12
/ Sequence 12, Application US/08765381
/ Patent No. 6083724
/ GENERAL INFORMATION:
/ APPLICANT: Commonwealth Scientific and Industrial Research Organisation
/ TITLE OF INVENTION: No. 6083724el avian cytokines and genetic
/ NUMBER OF SEQUENCES: 17
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Scully Scott Murphy and Presser
/ STREET: 400 Garden City Plaza
/ CITY: Garden City, New York
/ STATE: New York
/ COUNTRY: UNITED STATES OF AMERICA
/ ZIP: 11530
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/765,381
/ FILING DATE: 19-DEC-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: AU PN1542/95
/ FILING DATE: 06-MAR-1995
/ PRIOR APPLICATION DATA: PCT/AU96/00114
/ FILING DATE: 05-MAR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: PRESSER, LEOPOLD
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 1-516-742-4343
/ TELEFAX: 1-516-742-4366
/ INFORMATION FOR SEQ ID NO: 12:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 166 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: NO
/ ORIGINAL SOURCE:
/ ORGANISM: ovine
US-08-765-381-12

Query Match          56.7%; Score 17; DB 3; Length 166;
Best Local Similarity 50.0%; Pred. No. 8.5e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 3 YTS SFL 8

RESULT 14
US-09-679-397-1
/ Sequence 1, Application US/09679397
/ Patent No. 6339142

/ GENERAL INFORMATION:
/ APPLICANT: BASEY, CAROL D.
/ TITLE OF INVENTION: PROTEIN PURIFICATION
/ FILE REFERENCE: P1241R1D2
/ CURRENT APPLICATION NUMBER: US/09/679,397
/ CURRENT FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: US 60/084,459
/ PRIOR FILING DATE: 1998-05-06
/ PRIOR APPLICATION NUMBER: US 09/304,465
/ PRIOR FILING DATE: 1999-05-03
/ NUMBER OF SEQ ID NOS: 2
/ SEQ ID NO 1
/ LENGTH: 214
/ TYPE: PRT
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Sequence is synthesized.
US-09-679-397-1

Query Match          56.7%; Score 17; DB 4; Length 214;
Best Local Similarity 50.0%; Pred. No. 1.1e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 49 YSASF L 54

RESULT 15
US-09-680-148-1
/ Sequence 1, Application US/09680148
/ Patent No. 6417335
/ GENERAL INFORMATION:
/ APPLICANT: BASEY, CAROL D.
/ TITLE OF INVENTION: PROTEIN PURIFICATION
/ FILE REFERENCE: P1241R1D1
/ CURRENT APPLICATION NUMBER: US/09/680,148
/ CURRENT FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: US 60/084,459
/ PRIOR FILING DATE: 1998-05-06
/ PRIOR APPLICATION NUMBER: US 09/304,465
/ PRIOR FILING DATE: 1999-05-03
/ NUMBER OF SEQ ID NOS: 2
/ SEQ ID NO 1
/ LENGTH: 214
/ TYPE: PRT
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Sequence is synthesized.
US-09-680-148-1

Query Match          56.7%; Score 17; DB 4; Length 214;
Best Local Similarity 50.0%; Pred. No. 1.1e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 49 YSASF L 54

Search completed: October 5, 2004, 16:15:41
Job time : 22.6986 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:12:48 ; Search time 77.1507 Seconds  
(without alignments)  
66.737 Million cell updates/sec

Title: US-09-973-473A-26  
Perfect score: 30  
Sequence: 1 XXXYXXFLXXXXXX 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	56.7	18	US-09-735-995-99	Sequence 99, Appl
2	17	56.7	18	US-10-696-708-99	Sequence 99, Appl
3	17	56.7	25	US-10-424-599-275914	Sequence 275914,
4	17	56.7	41	US-09-864-761-41815	Sequence 41815, A
5	17	56.7	44	US-10-424-599-210212	Sequence 210212,
6	17	56.7	49	US-10-424-599-167405	Sequence 167405,
7	17	56.7	63	US-10-424-599-253802	Sequence 253802,
8	17	56.7	67	US-10-424-599-161117	Sequence 161117,
9	17	56.7	68	US-10-424-599-152881	Sequence 152881,
10	17	56.7	68	US-10-424-599-182964	Sequence 182964,
11	17	56.7	72	US-10-424-599-222881	Sequence 222881,
12	17	56.7	82	US-10-424-599-223775	Sequence 223775,
13	17	56.7	85	US-10-424-599-240477	Sequence 240477,
14	17	56.7	91	US-10-424-599-163312	Sequence 163312,
15	17	56.7	95	US-10-424-599-183752	Sequence 183752,

16	17	56.7	98	12	US-10-424-599-156773	Sequence 156773,
17	17	56.7	103	12	US-10-424-599-160140	Sequence 160140,
18	17	56.7	103	16	US-10-379-392-102	Sequence 102, App
19	17	56.7	103	16	US-10-379-392-127	Sequence 127, App
20	17	56.7	103	16	US-10-379-392-133	Sequence 133, App
21	17	56.7	107	9	US-09-949-559-9	Sequence 9, Appli
22	17	56.7	107	9	US-09-949-559-11	Sequence 11, Appl
23	17	56.7	107	10	US-09-875-221A-9	Sequence 9, Appli
24	17	56.7	107	10	US-09-875-221A-11	Sequence 11, Appl
25	17	56.7	107	12	US-10-411-037-35	Sequence 35, Appl
26	17	56.7	107	12	US-10-411-037-35	Sequence 35, Appl
27	17	56.7	107	12	US-10-411-026-35	Sequence 29, Appl
28	17	56.7	107	15	US-10-412-703A-29	Sequence 29, Appl
29	17	56.7	107	16	US-10-410-962-35	Sequence 35, Appl
30	17	56.7	107	16	US-10-411-049-35	Sequence 35, Appl
31	17	56.7	107	16	US-10-410-930-35	Sequence 35, Appl
32	17	56.7	107	16	US-10-410-997-35	Sequence 35, Appl
33	17	56.7	107	16	US-10-411-012-35	Sequence 35, Appl
34	17	56.7	107	16	US-10-287-994-35	Sequence 35, Appl
35	17	56.7	107	16	US-10-410-913-35	Sequence 35, Appl
36	17	56.7	108	9	US-09-971-543-7	Sequence 7, Appli
37	17	56.7	108	9	US-09-949-559-104	Sequence 104, App
38	17	56.7	108	10	US-09-875-221A-104	Sequence 104, App
39	17	56.7	109	15	US-10-410-894-1	Sequence 1, Appli
40	17	56.7	109	15	US-10-379-392-149	Sequence 149, App
41	17	56.7	113	12	US-10-424-599-162432	Sequence 162432,
42	17	56.7	116	16	US-10-767-701-39496	Sequence 39496, A
43	17	56.7	119	12	US-10-282-122A-71293	Sequence 71293, A
44	17	56.7	134	16	US-10-767-701-56579	Sequence 56579, A
45	17	56.7	137	16	US-10-767-701-61457	Sequence 61457, A
	17	56.7	140	12	US-10-425-114-62109	Sequence 62109, A

ALIGNMENTS

RESULT 1

US-09-735-995-99  
; Sequence 99, Application US/09735995  
; Patent No. US20010034024A1  
; GENERAL INFORMATION:  
; APPLICANT: Keating, Mark T.  
; APPLICANT: Splawski, Igor  
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT  
; FILE REFERENCE: 2323-136  
; CURRENT APPLICATION NUMBER: US/09/735.995  
; CURRENT FILING DATE: 2000-12-14  
; PRIOR APPLICATION NUMBER: 09/226,012  
; PRIOR FILING DATE: 1999-01-06  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 99  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-735-995-99

Query Match 56.7%; Score 17; DB 9; Length 18;  
Best Local Similarity 50.0%; Pred. No. 4.4e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXYFL 9  
Db 10 YSNAFL 15

RESULT 2

US-10-696-708-99  
; Sequence 99, Application US/10696708  
; Publication No. US2004007883A1  
; GENERAL INFORMATION:  
; APPLICANT: Keating, Mark T.  
; APPLICANT: Splawski, Igor

;/ TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT

;/ FILE OF INVENTION: SYNDROME GENE  
;/ FILE REFERENCE: 2323-164  
;/ CURRENT APPLICATION NUMBER: US/10/696,708  
;/ CURRENT FILING DATE: 2003-10-30  
;/ PRIOR APPLICATION NUMBER: US 09/735,995  
;/ PRIOR FILING DATE: 2000-12-14  
;/ PRIOR APPLICATION NUMBER: US 09/226,012  
;/ PRIOR FILING DATE: 1999-01-06  
;/ PRIOR APPLICATION NUMBER: 09/122,847  
;/ PRIOR FILING DATE: 1998-07-27  
;/ NUMBER OF SEQ ID NOS: 116  
;/ SOFTWARE: PatentIn Ver. 2.0  
;/ SEQ ID NO 99  
;/ LENGTH: 18  
;/ TYPE: PRT  
;/ ORGANISM: Homo sapiens  
;/ US-10-696-708-99

Query Match 56.7%; Score 17; DB 16; Length 18;  
Best Local Similarity 50.0%; Pred. No. 4.4e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
| | | |  
Db 10 YSAAFL 15

RESULT 3  
US-10-424-599-275914  
;/ Sequence 275914, Application US/10424599  
;/ Publication No. US20040031072A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: La Rosa Thomas J  
;/ APPLICANT: Kovalic David K  
;/ APPLICANT: Zhou Yihua  
;/ APPLICANT: Cao Yongwei  
;/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
;/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
;/ FILE REFERENCE: 38-21(53223)B  
;/ CURRENT APPLICATION NUMBER: US/10/424,599  
;/ CURRENT FILING DATE: 2003-04-28  
;/ NUMBER OF SEQ ID NOS: 285684  
;/ SEQ ID NO 275914  
;/ LENGTH: 25  
;/ TYPE: PRT  
;/ ORGANISM: Glycine max  
;/ FEATURE:  
;/ OTHER INFORMATION: Clone ID: PAT\_MRT3847\_91170C.1.pep  
US-10-424-599-275914

Query Match 56.7%; Score 17; DB 12; Length 25;  
Best Local Similarity 50.0%; Pred. No. 5.9e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
| | | |  
Db 6 YAAFL 11

RESULT 4  
US-09-864-761-41815  
;/ Sequence 41815, Application US/09864761  
;/ Patent No. US20020048763A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Penn, Sharon G.  
;/ APPLICANT: Rank, David R.  
;/ APPLICANT: Hanzel, David K.  
;/ APPLICANT: Chen, Wensheng  
;/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;/ TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
;/ FILE REFERENCE: Aecomica-X-1  
;/ CURRENT APPLICATION NUMBER: US/09/864,761

;/ CURRENT FILING DATE: 2001-05-23  
;/ PRIOR APPLICATION NUMBER: US 60/180,312  
;/ PRIOR FILING DATE: 2000-02-04  
;/ PRIOR APPLICATION NUMBER: US 60/207,456  
;/ PRIOR FILING DATE: 2000-05-26  
;/ PRIOR APPLICATION NUMBER: US 09/632,366  
;/ PRIOR FILING DATE: 2000-08-03  
;/ PRIOR APPLICATION NUMBER: GB 24263.6  
;/ PRIOR FILING DATE: 2000-10-04  
;/ PRIOR APPLICATION NUMBER: US 60/236,359  
;/ PRIOR FILING DATE: 2000-09-27  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00666  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00667  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00664  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00669  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00665  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00668  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00663  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00662  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00661  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: PCT/US01/00670  
;/ PRIOR FILING DATE: 2001-01-30  
;/ PRIOR APPLICATION NUMBER: US 60/234,687  
;/ PRIOR FILING DATE: 2000-09-21  
;/ PRIOR APPLICATION NUMBER: US 09/608,408  
;/ PRIOR FILING DATE: 2000-08-30  
;/ PRIOR APPLICATION NUMBER: US 09/774,203  
;/ PRIOR FILING DATE: 2001-01-29  
;/ NUMBER OF SEQ ID NOS: 49117  
;/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;/ SEQ ID NO 41815  
;/ LENGTH: 41  
;/ TYPE: PRT  
;/ ORGANISM: Homo sapiens  
;/ FEATURE:  
;/ OTHER INFORMATION: MAP TO AC018720.2  
;/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7  
;/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6  
;/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 9.7  
;/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2  
;/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.6  
;/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.9  
;/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.5  
;/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.9  
;/ OTHER INFORMATION: EST HUMAN HIT: BE388493.1, EVALUE 9.00e-19  
;/ OTHER INFORMATION: SWISSPROT HIT: O13816, EVALUE 2.00e-06  
US-09-864-761-41815

Query Match 56.7%; Score 17; DB 9; Length 41;  
Best Local Similarity 50.0%; Pred. No. 9.1e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
| | | |  
Db 21 YSTSFL 26

RESULT 5  
US-10-424-599-210212  
;/ Sequence 210212, Application US/10424599  
;/ Publication No. US20040031072A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: La Rosa Thomas J  
;/ APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 210212  
; LENGTH: 44  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_3184C.1.pep  
US-10-424-599-210212

Query Match 56.7%; Score 17; DB 12; Length 44;  
Best Local Similarity 50.0%; Pred. No. 9.7e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXFL 9  
| | |  
Db 11 YSSSFL 16

## RESULT 6

US-10-424-599-167405  
; Sequence 167405, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 167405  
; LENGTH: 49  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_122184C.1.pep  
US-10-424-599-167405

Query Match 56.7%; Score 17; DB 12; Length 49;  
Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXFL 9  
| | |  
Db 26 YTSFL 31

## RESULT 7

US-10-424-599-253802  
; Sequence 253802, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 253802

; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_71206C.1.pep  
US-10-424-599-253802

Query Match 56.7%; Score 17; DB 12; Length 63;  
Best Local Similarity 50.0%; Pred. No. 1.3e+03;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXFL 9  
| | |  
Db 24 YSSTFL 29

## RESULT 8

US-10-424-599-161117  
; Sequence 161117, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 161117  
; LENGTH: 67  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_116506C.1.pep  
US-10-424-599-161117

Query Match 56.7%; Score 17; DB 12; Length 67;  
Best Local Similarity 50.0%; Pred. No. 1.4e+03;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXFL 9  
| | |  
Db 12 YATTFL 17

## RESULT 9

US-10-424-599-152881  
; Sequence 152881, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 152881  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_109075C.1.pep  
US-10-424-599-152881

Query Match 56.7%; Score 17; DB 12; Length 68;  
Best Local Similarity 50.0%; Pred. No. 1.4e+03;

```
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9
Db 40 YSATFL 45

RESULT 10
US-10-424-599-182964
; Sequence 182964, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 182964
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(68)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_136222C.1.pep
US-10-424-599-182964

Query Match 56.7%; Score 17; DB 12; Length 68;
Best Local Similarity 50.0%; Pred. No. 1.4e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9
Db 2 YTTFL 7

RESULT 11
US-10-424-599-222881
; Sequence 222881, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 222881
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_43282C.1.pep
US-10-424-599-222881

Query Match 56.7%; Score 17; DB 12; Length 72;
Best Local Similarity 50.0%; Pred. No. 1.5e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9
Db 64 YSASFL 69

RESULT 12
US-10-424-599-223775
; Sequence 223775, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223775
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44099C.1.pep
US-10-424-599-223775

Query Match 56.7%; Score 17; DB 12; Length 82;
Best Local Similarity 50.0%; Pred. No. 1.7e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9
Db 28 YSTFL 33

RESULT 13
US-10-424-599-240477
; Sequence 240477, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 240477
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_59179C.1.pep
US-10-424-599-240477

Query Match 56.7%; Score 17; DB 12; Length 85;
Best Local Similarity 50.0%; Pred. No. 1.7e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9
Db 49 YSSFL 54

RESULT 14
US-10-424-599-163312
; Sequence 163312, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
```

```
QY 4 YXXFL 9
Db 64 YSASFL 69
```

; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 163312  
; LENGTH: 91  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(91)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_11848C.1.pep  
US-10-424-599-163312

Query Match 56.7%; Score 17; DB 12; Length 91;  
Best Local Similarity 50.0%; Pred. No. 1.8e+03;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 YXXFL 9  
| | |  
Db 61 YAAFL 66

RESULT 15  
US-10-424-599-183752  
; Sequence 183752, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 183752  
; LENGTH: 95  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_136942C.1.pep  
US-10-424-599-183752

Query Match 56.7%; Score 17; DB 12; Length 95;  
Best Local Similarity 50.0%; Pred. No. 1.9e+03;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 YXXFL 9  
| | |  
Db 48 YTSFL 53

Search completed: October 5, 2004, 16:47:08  
Job time : 78.1507 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:58:01 ; Search time 19.5068 Seconds  
(without alignments)  
78.899 Million cell updates/sec

Title: US-09-973-473A-26  
Perfect score: 30  
Sequence: 1 XXXYXXFLXXXXXX 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	56.7	38	2 S22210	photosystem I prot
2	17	56.7	70	2 I77547	hypothetical prote
3	17	56.7	112	2 S58139	gene 9 protein - p
4	17	56.7	141	2 C82612	hypothetical prote
5	17	56.7	154	2 F70902	hypothetical prote
6	17	56.7	156	2 A86979	probable membrane
7	17	56.7	166	2 JC6559	interferon-gamma p
8	17	56.7	166	2 S12723	interferon-gamma p
9	17	56.7	172	2 T25500	hypothetical prote
10	17	56.7	193	1 S07734	NADH2 dehydrogenas
11	17	56.7	219	2 F65018	hypothetical prote
12	17	56.7	219	2 F91042	hypothetical prote
13	17	56.7	219	2 A85887	hypothetical prote
14	17	56.7	219	2 AE0813	ethanolamine utili
15	17	56.7	223	2 T21930	hypothetical prote
16	17	56.7	228	2 G69372	hypothetical prote
17	17	56.7	237	2 A43984	virulence-associat
18	17	56.7	241	1 S08402	virulence associat
19	17	56.7	241	2 A41481	virulence associat
20	17	56.7	241	2 JQ0747	virulence-associat
21	17	56.7	245	2 C90540	conserved hypothet
22	17	56.7	246	2 T47398	hypothetical prote
23	17	56.7	250	2 A41416	hypothetical prote
24	17	56.7	260	2 G84960	enoyl-facyl-carrie
25	17	56.7	281	2 T13596	trypsin homolog -
26	17	56.7	286	2 C85061	probable phosphati
27	17	56.7	295	2 E69992	conserved hypothet
28	17	56.7	295	2 T42515	hypothetical prote
29	17	56.7	298	2 A82879	conserved hypothet

30	17	56.7	303	2 E86591	CT580 hypothetical
31	17	56.7	303	2 D72031	ct580 hypothetical
32	17	56.7	311	2 B70083	hypothetical prote
33	17	56.7	325	2 AB1123	B. subtilis tolC p
34	17	56.7	325	2 AD1483	B. subtilis tolC p
35	17	56.7	331	2 G83939	myo-inositol catab
36	17	56.7	374	2 T09106	gibberellin 20-oxi
37	17	56.7	379	2 T24654	hypothetical prote
38	17	56.7	386	2 S60646	NADH2 dehydrogenas
39	17	56.7	392	2 T25209	hypothetical prote
40	17	56.7	399	2 T18853	probable transcrip
41	17	56.7	404	2 T40675	hypothetical prote
42	17	56.7	408	2 T08400	late embryonic abu
43	17	56.7	431	2 T07812	S-locus-specific g
44	17	56.7	443	2 S37612	NADH2 dehydrogenas
45	17	56.7	445	2 E87561	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S22210 photosystem I protein psal - Synechococcus sp.

N:Alternate names: Photosystem I chain VIII

C:Species: Synechococcus sp.

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Mar-2001

C:Accession: S22210

R:Muehlenhoff, U.; Haehnel, W.; Witt, H.T.; Herrmann, R.G.

submitted to the EMBL Data Library, January 1992

A:Description: Genes encoding ten subunits of photosystem I

A:Reference number: S18970

A:Accession: S22210

A:Molecule type: DNA

A:Residues: 1-38 <MUE>

A:Cross-references: EMBL:X63763; NID:G47585; PIDN:CAA45297.1; PID:G47588

C:Genetics:

A:Gene: psal

C:Superfamily: photosystem I protein psal

C:Keywords: membrane-associated complex; photosynthesis; photosystem I; thylakoid

Query Match 56.7%; Score 17; DB 2; Length 38;

Best Local Similarity 50.0%; Fred. No. 69;

Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXPL 9

Db 5 YAASFL 10

##### RESULT 2

I77547

hypothetical protein 2 - Escherichia coli insertion sequence IS903

C:Species: Escherichia coli

C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 20-Sep-1999

C:Accession: I77547

R:Mollet, B.; Iida, S.; Arber, W.

Mol. Gen. Genet. 199, 534-536, 1985

A:Title: An active variant of the prokaryotic transposable element IS903 carries an amber mutation in the coding region

A:Reference number: I57738; MUID:85295477; PMID:2993802

A:Accession: I77547

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-70 <RES>

A:Cross-references: EMBL:X02527; NID:G43694; PIDN:CAA26363.1; PID:G43696

A:Experimental source: strain K-12, subspecies WA921

C:Genetics:

A:Mobile element: insertion sequence IS903

C:Superfamily: hypothetical protein IR903

Query Match 56.7%; Score 17; DB 2; Length 70;

Best Local Similarity 50.0%; Fred. No. 1.2e+02;

Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
 |  
 Db 14 YTAFL 19  
 |  
 RESULT 3  
 S58139  
 gene 9 protein - phage SPPI  
 C:Species: phage SPPI  
 C>Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 11-May-2000  
 C:Accession: S58139; T42277  
 R:Becker, B.; Gassel, M.; Tavares, P.; Lurz, R.; Alonso, J.C.  
 submitted to the EMBL Data Library, July 1995  
 A:Description: Head morphogenesis of the Bacillus subtilis bacteriophage SPPI.  
 A:Reference number: S58137  
 A:Accession: S58139  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-112 <BEC>  
 A:Cross-references: EMBL:X89721; NID:g1052805; PIDN:CAA61867.1; PID:g1052808  
 R:Alonso, J.C.; Luder, G.; Stiege, A.C.; Chai, S.; Weise, F.; Trautner, T.A.  
 Gene 204, 201-212, 1997  
 A:Title: The complete nucleotide sequence and functional organization of Bacillus subtilis  
 A:Reference number: Z22137; MUID:98094274; PMID:9434185  
 A:Accession: T42277  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-112 <ALO>  
 A:Cross-references: EMBL:X97918; PIDN:CAA66587.1  
 C:Superfamily: phage SPPI gene 9 protein  
 Query Match 56.7%; Score 17; DB 2; Length 112;  
 Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 YXXXFL 9  
 |  
 Db 105 YTSFL 110  
 |  
 RESULT 4  
 C82612  
 hypothetical protein XF2001 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: C82612  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A>Note: for a complete list of authors see reference number A59328 below  
 A:Accession: C82612  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-141 <SIM>  
 A:Cross-references: GB:AE004019; GB:AE003849; NID:g9107105; PIDN:AAF84803.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 br-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Kriesger, J.E.; Kuranae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.P.; Marino, C.L.; Marques, M.V.; Martins  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:

A:Gene: XF2001  
 Query Match 56.7%; Score 17; DB 2; Length 141;  
 Best Local Similarity 50.0%; Pred. No. 2.4e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 YXXXFL 9  
 |  
 Db 128 YAAAF 133  
 |  
 RESULT 5  
 F70902  
 hypothetical protein Rv1417 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 28-Jul-2000  
 C:Accession: F70902  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.;  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: F70902  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-154 <COL>  
 A:Cross-references: GB:Z80108; GB:AL123456; NID:g3256012; PIDN:CAB02165.1; PID:e265578; I  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: Rv1417  
 C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1417  
 Query Match 56.7%; Score 17; DB 2; Length 154;  
 Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 YXXXFL 9  
 |  
 Db 22 YAAAF 27  
 |  
 RESULT 6  
 A86979  
 probable membrane protein [imported] - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C:Accession: A86979  
 R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor  
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holtroyd,  
 eam, M.A.; Rutherford, K.M.  
 Nature 409, 1007-1011, 2001  
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
 A:Title: Massive gene decay in the leprosy bacillus.  
 A:Reference number: A86909; MUID:21128732; PMID:11234002  
 A:Accession: A86979  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-156 <STO>  
 A:Cross-references: GB:AL450380; NID:gl3092760; PIDN:CAC30069.1; GSPDB:GN00147  
 C:Genetics:  
 A:Gene: ML0561  
 C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1417  
 Query Match 56.7%; Score 17; DB 2; Length 156;  
 Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 YXXXFL 9  
 |  
 Db 24 YAAAF 29  
 |



## RESULT 7

JC6559  
interferon-gamma precursor - wild goat  
C:Species: Capra aegagrus (wild goat)  
C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 07-May-1999  
C:Accession: JC6559  
R:Beyer, J.C.; Stich, R.W.; Hoover, D.S.; Brown, W.C.; Chéevers, W.P.  
Gene 210, 103-108, 1998  
A:Title: Cloning and expression of caprine interferon-gamma.  
A:Reference number: JC6559; MUID:98192545; PMID:9524237  
A:Accession: JC6559  
A:Molecule type: mRNA  
A:Residues: 1-166 <BEY>  
A:Cross-references: GB:U34232  
C:Superfamily: interferon gamma  
C:Keywords: cytokine; glycoprotein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-166/Product: interferon-gamma #status predicted <MAT>  
F:739,106/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 56.7%; Score 17; DB 2; Length 166;  
Best Local Similarity 50.0%; Pred. No. 2.8e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXFL 9  
| | |  
Db 3 YTSFSL 8

## RESULT 8

S12723  
interferon gamma precursor - sheep  
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
C:Accession: S12723  
R:McInnes, C.J.; Logan, M.; Redmond, J.; Entrican, G.; Baird, G.D.  
Nucleic Acids Res. 18, 4012, 1990  
A:Title: The molecular cloning of the ovine gamma-interferon cDNA using the polymerase  
A:Reference number: S12723; MUID:90326548; PMID:2115673  
A:Accession: S12723  
A:Molecule type: mRNA  
A:Residues: 1-166 <MCI>  
A:Cross-references: EMBL:X52640; NID:g1796; PIDN:CAA36862.1; PID:g1797  
C:Superfamily: interferon gamma

Query Match 56.7%; Score 17; DB 2; Length 166;  
Best Local Similarity 50.0%; Pred. No. 2.8e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXFL 9  
| | |  
Db 3 YTSFSL 8

## RESULT 9

T25500  
hypothetical protein C03G6.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25500  
R:Murray, J.; Wohldmann, P.  
submitted to the EMBL Data Library, April 1997  
A:Description: The sequence of C. elegans cosmid C03G6.  
A:Reference number: Z20042  
A:Accession: T25500  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-172 <MUR>  
A:Cross-references: EMBL:U97008; PIDN:AB52301.1; GSPDB:GN00023; CESP:C03G6.1  
A:Experimental source: strain Bristol N2; clone C03G6  
C:Genetics:  
A:Gene: CESP:C03G6.1

A:Map position: 5  
A:Introns: 5/3; 59/2; 133/1

Query Match 56.7%; Score 17; DB 2; Length 172;  
Best Local Similarity 50.0%; Pred. No. 2.9e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXFL 9  
| | |  
Db 45 YSSTFL 50

## RESULT 10

S07734  
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - Paramesidium tetraurelia mitochondrion  
A:Alternate names: NADH-ubiquinone oxidoreductase chain 2  
C:Species: mitochondrion Paramesidium tetraurelia  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Jun-2002  
C:Accession: S07734; JS0233  
R:Pritchard, A.E.; Seilhamer, J.J.; Mahalingam, R.; Sable, C.L.; Venuti, S.E.; Cummings, Nucleic Acids Res. 18, 173-180, 1990  
A:Title: Nucleotide sequence of the mitochondrial genome of Paramesidium.  
A:Reference number: S07725; MUID:90174913; PMID:2308823  
A:Accession: S07734  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-193 <PRI1>  
A:Cross-references: EMBL:X15917; NID:g13256; PIDN:CAA34043.1; PID:g515876  
R:Pritchard, A.E.; Venuti, S.E.; Ghalambor, M.A.; Sable, C.L.; Cummings, D.J. Gene 78, 121-134, 1989  
A:Title: An unusual region of Paramesidium mitochondrial DNA containing chloroplast-like genes  
A:Reference number: JS0231; MUID:89357489; PMID:2670676  
A:Accession: JS0233  
A:Molecule type: DNA  
A:Residues: 'L', 2-193 <PRI2>  
A:Cross-references: GB:M26930; NID:g341550; PIDN:AAA79255.1; PID:g1019630  
A:Experimental source: strain sp. 4.51  
A:Note: The authors translated the initiation codon TTG for residue 1 as Leu  
C:Genetics:  
A:Gene: ndh2  
A:Genome: mitochondrion  
A:Genetic code: SGC6  
A:Start codon: TTG  
C:Superfamily: Paramesidium NADH dehydrogenase (ubiquinone) chain 2  
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; c

Query Match 56.7%; Score 17; DB 1; Length 193;  
Best Local Similarity 50.0%; Pred. No. 3.2e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 YXXXFL 9  
| | |  
Db 19 YTSFSL 24

## RESULT 11

F65018  
hypothetical protein b2439 - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 14-Jul-2003  
C:Accession: F65018  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coli Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F65018  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-219 <BLAT>  
A:Cross-references: GB:AB000331; GB:U00096; NID:g1788775; PIDN:AA75492.1; PID:g1788780;  
A:Experimental source: strain K-12, substrain MGL655  
C:Superfamily: polyhedral organelle shell protein, EutL/PduB type

```

Query Match          56.7%; Score 17; DB 2; Length 219;
Best Local Similarity 50.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      |  |  |  |
Db      185 YSAAFL 190

RESULT 12
F91042
hypothetical protein ECs3310 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Jul-2003
C:Accession: F91042
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F91042
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-219 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA836733.1; PID:g13362780; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs3310
C:Superfamily: polyhedral organelle shell protein, EutL/PduB type

Query Match          56.7%; Score 17; DB 2; Length 219;
Best Local Similarity 50.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      |  |  |  |
Db      185 YSAAFL 190

RESULT 13
A85887
hypothetical protein Z3704 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Jul-2003
C:Accession: A85887
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85887
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-219 <STO>
A:Cross-references: GB:AE005174; NID:g12516820; PIDN:AAG57557.1; GSPDB:GN00145; UWGP:Z37
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3704
C:Superfamily: polyhedral organelle shell protein, EutL/PduB type

Query Match          56.7%; Score 17; DB 2; Length 219;
Best Local Similarity 50.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      |  |  |  |
Db      185 YSAAFL 190

RESULT 14
AE0813
ethanolamine utilization protein EutL [imported] - Salmonella enterica subsp. enterica s

```

```

C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 14-Jul-2003
C:Accession: AE0813
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conington, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0813
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-219 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07687.1; PID:g16503673; GSPDB:GN00176
C:Genetics:
A:Gene: eutL
C:Superfamily: polyhedral organelle shell protein, EutL/PduB type

Query Match          56.7%; Score 17; DB 2; Length 219;
Best Local Similarity 50.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      |  |  |  |
Db      185 YSAAFL 190

RESULT 15
T21930
hypothetical protein F38A1.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21930
R:DiIoyd, C.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19489
A:Accession: T21930
A>Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-223 <WIL>
A:Cross-references: EMBL:Z81535; PIDN:CAB04353.1; GSPDB:GN00022; CESP:F38A1.1
A:Experimental source: clone F38A1
C:Genetics:
A:Gene: CESP:F38A1.1
A:Map position: 4
A:Introns: 73/1; 193/1

Query Match          56.7%; Score 17; DB 2; Length 223;
Best Local Similarity 50.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      |  |  |  |
Db      66 YTSSFL 71

Search completed: October 5, 2004, 16:13:59
Job time : 21.5068 secs

```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:36 ; Search time 11.1781 Seconds  
(without alignments)  
74.532 Million cell updates/sec

Title: US-09-973-473A-26  
Perfect score: 30  
Sequence: 1 XXXYXXFLXXXXXX 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	17	56.7	38	1	PSAI_SYNEL
2	17	56.7	38	1	PSAI_SYNP2
3	17	56.7	154	1	YE17_MYCTU
4	17	56.7	162	1	KCH2_CAVPO
5	17	56.7	166	1	ING_CAPHI
6	17	56.7	166	1	ING_SHEEP
7	17	56.7	167	1	VHR2_YABAM
8	17	56.7	178	1	DUSP_MYXVL
9	17	56.7	183	1	SFP4_BOVIN
10	17	56.7	193	1	NU2M_PARTE
11	17	56.7	219	1	EUTL_ECOLI
12	17	56.7	219	1	EUTL_SALTY
13	17	56.7	236	1	VRP4_SALTY
14	17	56.7	240	1	VRP3_SALCH
15	17	56.7	240	1	VRP3_SALTY
16	17	56.7	260	1	FABI_BUCAI
17	17	56.7	295	1	MNTD_BACSU
18	17	56.7	298	1	Y535_UREPA
19	17	56.7	311	1	YXAF_BACSU
20	17	56.7	379	1	GRR2_CERAE
21	17	56.7	379	1	GRR2_HUMAN
22	17	56.7	379	1	GRR2_MACMU
23	17	56.7	380	1	GRR2_CALJA
24	17	56.7	386	1	NUAM_ARTSF
25	17	56.7	414	1	R3R2_MOUSE
26	17	56.7	436	1	TX6_HUMAN
27	17	56.7	443	1	NUAM_CHLRE
28	17	56.7	470	1	TACY_STRPN
29	17	56.7	478	1	MOT2_HUMAN
30	17	56.7	485	1	TACY_BACCE
31	17	56.7	485	1	Y44E_CABEL
32	17	56.7	489	1	C302_DROME
33	17	56.7	526	1	KCH2_CHICK

#### ALIGNMENTS

##### RESULT 1

```

PSAI_SYNEL
ID   PSAI_SYNEL          STANDARD;          PRT;          38 AA.
AC   P25900;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Photosystem I reaction center subunit VIII.
GN   PSAI OR TSR2405.
OS   Synechococcus elongatus (Thermosynechococcus elongatus), and
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=32046, 1141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.elongatus; STRAIN=BP-1;
RX   MEDLINE=2225144; PubMed=12240834;
RA   Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA   Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the thermophilic cyanobacterium
RT   Thermosynechococcus elongatus BP-1.";
RL   DNA Res. 9:123-130(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.e.naegeli;
RX   MEDLINE=9325282; PubMed=8486290;
RA   Muehlenhoff U., Haehnel W., Witt H.T., Herrmann R.G.;
RT   "Genes encoding eleven subunits of photosystem I from the
RT   thermophilic cyanobacterium Synechococcus sp.";
RL   Gene 127:71-78(1993).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (4.0 ANGSTROMS).
RC   SPECIES=S.e.naegeli;
RX   MEDLINE=97057537; PubMed=8901876;
RA   Krauss N., Schubert W.D., Klukas O., Fromme P., Witt H.T., Saenger W.;
RT   "Photosystem I at 4-A resolution represents the first structural
RT   model of a joint photosynthetic reaction centre and core antenna
RT   system.";
RL   Nat. Struct. Biol. 3:965-973(1996).
CC   -!- FUNCTION: May help in the organization of the psal subunit.
CC   -!- SIMILARITY: Belongs to the psal family.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; AP005377; BAC09957.1; -.
DR   EMBL; X63763; CAA45297.1; -.
DR   PDB; 2PPS; 27-MAY-98.
DR   HAMAP; MF_00431; -; 1.

```

```

34 17 56.7 663 1 TKT1_ECOLI
35 17 56.7 665 1 TKT_BUCAI
36 17 56.7 665 1 TKT_BUCAP
37 17 56.7 667 1 TKT2_ECOLI
38 17 56.7 745 1 FSP1_RAT
39 17 56.7 912 1 MGR4_HUMAN
40 17 56.7 912 1 MGR4_RAT
41 17 56.7 950 1 KCH6_RAT
42 17 56.7 994 1 KCH6_HUMAN
43 17 56.7 1158 1 KCH2_CANFA
44 17 56.7 1159 1 KCH2_HUMAN
45 17 56.7 1161 1 KCH2_RABIT

```

```

P27302 escherichia
P57195 buchnera ap
Q8ka26 buchnera ap
P33570 escherichia
Q63517 rattus norv
Q14833 homo sapien
P14823 rattus norv
P31423 rattus norv
O54853 rattus norv
Q9t252 homo sapien
Q9t253 canis famil
Q12809 homo sapien
Q8wny2 oryctolagus

```

```

DR InterPro; IPR001302; PSI 8.
DR Pfam; PF00796; PSI 8; 1.
DR ProDom; PD003995; PSI 8; 1.
KW Photosystem I; Photosynthesis; Transmembrane; 3D-structure;
KW Complete proteome. 32 POTENTIAL.
FT TRANSMEM 12 32
SQ SEQUENCE 38 AA; 4297 MW; E9B0178560DE5CF5 CRC64;

Query Match 56.7%; Score 17; DB 1; Length 38;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
| |||
Db 5 YAAAF 10

RESULT 2
PSAI_SYN2 STANDARD; PRT; 38 AA.
AC Q54752;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem I reaction center subunit VIII.
GN PSAI.
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32049;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96271772; PubMed=8787020;
RA Schlucher W.M., Shen G., Zhao J., Bryant D.A.;
RT "Characterization of psal and psal mutants of Synechococcus sp.
RT strain PCC 7002: a new model for state transitions in
RT cyanobacteria.";
RL Photochem. Photobiol. 64:53-66(1996).
CC -!- FUNCTION: May help in the organization of the psal subunit.
CC -!- SIMILARITY: Belongs to the psal family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U58035; AAB18909.1; -.
DR HAMAP; MF_00431; -.
DR InterPro; IPR001302; PSI 8.
DR Pfam; PF00796; PSI 8; 1.
KW Photosystem I; Photosynthesis; Transmembrane.
FT TRANSMEM 9 29
SQ SEQUENCE 38 AA; 3957 MW; BE58C30EFAB31832 CRC64;

Query Match 56.7%; Score 17; DB 1; Length 38;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
| |||
Db 5 YAAAF 10

RESULT 3
YE17_MYCTU STANDARD; PRT; 154 AA.
AC P71686;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein Rv1417/MT1460/Mb1452.

```

---

```

GN RV1417 OR MT1460 OR MTCY21B4.35 OR MBL1452.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; Z80108; CAB02165.1; -.
DR EMBL; AE007016; AAK45725.1; -.
DR EMBL; BX248338; CAD94313.1; -.
DR FIR; F70902; F70902.
DR TIGR; MT1460; -.
DR Tuberculist; Rv1417; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 19 39
FT TRANSMEM 51 71
SQ SEQUENCE 154 AA; 16383 MW; 3109AB9ABDD3296D CRC64;

Query Match 56.7%; Score 17; DB 1; Length 154;
Best Local Similarity 50.0%; Pred. No. 13e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
| |||
Db 22 YAAAF 27

RESULT 4

```

```

KCH2_CAVPO
ID_KCH2_CAVPO STANDARD; PRT; 162 AA.
AC O08703;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Potassium voltage-gated channel subfamily H member 2 (Ether-a-go-go
DE related gene potassium channel 1) (ERG1) (gp-erg) (Ether-a-go-go
DE related protein 1) (Eag related protein 1) (Fragment).
GN KCNH2 OR ERG.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=97164986; PubMed=9012748;
RA Wymore R.S., Gintant G.A., Wymore R.T., Dixon J.E., McKinnon D.,
RA Cohen I.S.;
RT "Tissue and species distribution of mRNA for the Ikr-like K+ channel,
RT erg.";
RL Circ. Res. 80:261-268(1997).
CC -!- FUNCTION: Pore-forming (alpha) subunit of voltage-gated inwardly
CC rectifying potassium channel. Channel properties are modulated by
CC cAMP and subunit assembly. Mediates the rapidly activating
CC component of the delayed rectifying potassium current in heart
CC (Ikr) (By similarity).
CC -!- SUBUNIT: The potassium channel is probably composed of a homo- or
CC heterotetrameric complex of pore-forming alpha subunits that can
CC associate with modulating beta subunits. Heteromultimer with
CC KCNH6/ERG2, KCNH7/ERG3, KCNE1 and KCNE2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Highly expressed in heart and brain.
CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is
CC characterized by a series of positively charged amino acids at
CC every third position.
CC -!- SIMILARITY: Belongs to the potassium channel family. H (Eag)
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75211; AAC53159.1; -.
CC InterPro; IPR005821; Ion trans.
CC InterPro; IPR005820; M+channel_nlg.
CC Pfam; PF00520; Ion trans; 1.
CC Transport; Ion transport; 1.
CC Potassium channel; Potassium; Voltage-gated channel;
CC Multigene family.
CC NON TER 1 1
CC TRANSMEM <1 17 SEGMENT S1 (POTENTIAL).
CC TRANSMEM 44 64 SEGMENT S2 (POTENTIAL).
CC DOMAIN 65 88 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 89 109 SEGMENT S3 (POTENTIAL).
CC TRANSMEM 114 134 SEGMENT S4 (POTENTIAL).
CC DOMAIN 135 162 CYTOPLASMIC (POTENTIAL).
CC NON TER 162 162
CC SEQUENCE 162 AA; 18382 MW; B94BF06DFC29729B CRC64;

Query Match 56.7%; Score 17; DB 1; Length 162;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 YXXXFL 9
Db 20 YSAAPL 25

RESULT 6
ING_SHEEP STANDARD; PRT; 166 AA.
AC P1773;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Interferon gamma precursor (IFN-gamma).
GN IFNG.
OS Eukaryotes (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

RESULT 5
ING_CAPHI STANDARD; PRT; 166 AA.
AC P79154;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Interferon gamma precursor (IFN-gamma).
GN IFNG.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98192545; PubMed=9524237;
RA Beyer J.C., Stich R.W., Hoover D.S., Brown W.C., Cheevers W.P.;
RT "Cloning and expression of caprine interferon-gamma.";
RL Gene 210:103-108(1998).
CC -!- FUNCTION: Produced by lymphocytes activated by specific antigens
CC or mitogens. IFN-gamma, in addition to having antiviral activity,
CC has important immunoregulatory functions. It is a potent activator
CC of macrophages, it has antiproliferative effects on transformed
CC cells and it can potentiate the antiviral and antitumor effects of
CC the type I interferons (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Released primarily from activated T
CC lymphocytes (By similarity).
CC -!- SIMILARITY: Belongs to the type II (or gamma) interferon family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U34232; AAB38525.1; -.
CC HSP; P07353; 1D9G.
CC InterPro; IPR002069; IFN-gamma.
CC Pfam; PF00714; IFN-gamma; 1.
CC ProDom; PD002435; IFN-gamma; 1.
CC CytoKine; Antiviral; Growth regulation; Glycoprotein; Signal.
CC SIGNAL 1 20 BY SIMILARITY.
CC CHAIN 21 166 INTERFERON GAMMA.
CC CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 166 AA; 19325 MW; 9381PE0EA0605D1B CRC64;

Query Match 56.7%; Score 17; DB 1; Length 166;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 YXXXFL 9
Db 3 YTSFLL 8

```

```
CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
CC Bovidae; Caprinae; Ovis.
CC NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90326548; PubMed=2115673;
RA McInnes C.J., Logan M., Redmond J., Entrican G., Baird G.D.;
RT "The molecular cloning of the ovine gamma-interferon cDNA using the
RT polymerase chain reaction.";
RL Nucleic Acids Res. 18:4012-4012(1990).
CC -!- FUNCTION: Produced by lymphocytes activated by specific antigens
CC or mitogens. IFN-gamma, in addition to having antiviral activity,
CC has important immunoregulatory functions. It is a potent activator
CC of macrophages, it has antiproliferative effects on transformed
CC cells and it can potentiate the antiviral and antitumor effects of
CC the type I interferons.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Released primarily from activated T
CC lymphocytes.
CC -!- SIMILARITY: Belongs to the type II (or gamma) interferon family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52640; CAA36862.1; -.
DR EMBL; A19173; CAA01450.1; -.
DR PIR; S12723; S12723.
DR HSSP; P07353; 1D96.
DR InterPro; IPR002069; IFN-gamma.
DR Pfam; PF00714; IFN-gamma; 1.
DR ProDom; PD002435; IFN-gamma; 1.
KW Cytokine; Antiviral; Growth regulation; Glycoprotein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 166 INTERFERON GAMMA.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
SQ SEQUENCE 166 AA; 19369 MW; 9232D31269805D0A CRC64;
Query Match 56.7%; Score 17; DB 1; Length 166;
Best Local Similarity 50.0%; Pred. No. 1.4e+02; Mismatches 0; Gaps 0;
Matches 3; Conservative 0; Indels 0;
QY 4 YXXXFL 9
DB 3 YTSFSL 8
RESULT 7
VHR2_YABAM STANDARD; PRT; 167 AA.
ID VHR2_YABAM
AC Q9QB4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Probable host range protein 2.
GN YAB.
OS Yaba monkey tumor virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Yatapoxvirus.
OX NCBI_TaxID=38804;
RN [1]
RP SEQUENCE FROM N.A.
RA Amano H., Morikawa S., Ueda Y., Miyamura T.;
RT "Nucleotide sequence of the central 50kbp region of Yaba virus DNA.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Plays a role for multiplication of the virus in
CC different cell types (By similarity).
```

```
CC -!- SIMILARITY: Belongs to the poxviruses C7 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015885; BAA88781.1; -.
DR InterPro; IPR004967; Pox_C7_F8A.
DR Pfam; PF03287; Pox_C7_F8A; 1.
DR PIRSF; PIRSF003779; VAC_C7L; 1.
SQ SEQUENCE 167 AA; 19700 MW; A16F77EB33E20497 CRC64;
Query Match 56.7%; Score 17; DB 1; Length 167;
Best Local Similarity 50.0%; Pred. No. 1.4e+02; Mismatches 0; Gaps 0;
Matches 3; Conservative 0; Indels 0;
QY 4 YXXXFL 9
DB 85 YSASFL 90
RESULT 8
DUSP_MXVVL STANDARD; PRT; 178 AA.
ID DUSP_MXVVL
AC Q85297; Q908N1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Dual specificity protein phosphatase (EC 3.1.3.48) (EC 3.1.3.16).
GN M069L OR ILL.
OS Myxoma virus (strain Lausanne).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Leporipoxvirus.
OX NCBI_TaxID=31530;
RN [1]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND MUTAGENESIS OF CYS-110.
RX MEDLINE=95133195; PubMed=7831813;
RA Mossman K., Ostergaard H., Upton C., McFadden G.;
RT "Myxoma virus and Shope fibroma virus encode dual-specificity
RT tyrosine/serine phosphatases which are essential for virus
RT viability.";
RL Virology 206:572-582(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20032073; PubMed=10562494;
RA Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.X.,
RA Macaulay C., Willer D., Evans D., McFadden G.;
RT "The complete DNA sequence of myxoma virus.";
RL Virology 284:298-318(1999).
CC -!- FUNCTION: Shows activity toward tyrosine-protein phosphate as well
CC as with serine-protein phosphate. It is an essential factor for
CC virus viability.
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -!- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC Non-receptor class dual specificity subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L31960; AAA66956.1; -.
DR EMBL; AF170726; AAP14957.1; -.
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR000387; TYR_phosphatase.
```

```

DR Pfam; PF00782; DSPC; 1.
DR SMART; SM00195; DSPC; 1.
DR PROSITE; PS00383; TYR PHOSPHATASE 1; 1.
DR PROSITE; PS0056; TYR PHOSPHATASE 2; 1.
DR PROSITE; PS0054; TYR PHOSPHATASE_DUAL; 1.
KW Hydrolase.
FT ACT_SITE 110 110 PHOSPHOCYCSTEINE INTERMEDIATE.
FT MUTAGEN 110 110 C->S: COMPLETE LOSS OF ACTIVITY.
FT CONFLICT 172 178 HKLKLFG -> T (IN REF. 1).
SQ SEQUENCE 178 AA; 20627 MW; 318C99B3400A885F CRC64;

Query Match
Best Local Similarity 56.7%; Score 17; DB 1; Length 178;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXHFL 9
Db 91 YATTFL 96

RESULT 9
SFP4 BOVIN
ID SFP4 BOVIN STANDARD; PRT; 183 AA.
AC P81019; O97868;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Seminal plasma protein BSP-30 kDa precursor (BSP-30K).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Salois D., Menard M., Paquette Y., Manjunath P.;
RT "Complete mRNA sequence of bovine seminal plasma 30K protein (BSP-30K).";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 26-183, AND CARBOHYDRATE-LINKAGE SITES.
RC TISSUE-Seminal plasma;
RX MEDLINE=97134580; PubMed=8980140;
RA Calvete J.J., Mann K., Sanz L., Rada M., Toepfer-Petersen E.;
RT "The primary structure of BSP-30K, a major lipid-, gelatin-, and
heparin-binding glycoprotein of bovine seminal plasma.";
RL FEBS Lett. 399:147-152 (1996).
CC -!- FUNCTION: BINDS TO SPERMATOZOA UPON EJACULATION AND MAY PLAY A
ROLE IN SPERM CAPACITATION. DISPLAYS HEPARIN-, GELATIN- AND
PHOSPHOLIPID-BINDING ACTIVITIES.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the seminal plasma protein family.
CC -!- SIMILARITY: Contains 2 fibronectin type II domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF057133; AAD17519.1; -.
DR HSSP; P02784; 1PDC.
DR InterPro; IPR000562; FN_Type_II.
DR Pfam; PF00040; fn2; 2.
DR PRINTS; PR00013; FNTYPEII.
DR ProDom; PD00095; FN2; 2.
DR SMART; SM00059; FN2; 2.
DR PROSITE; PS00023; FIBRONECTIN 2; 1.
KW Fertilization; Glycoprotein; Heparin-binding; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 183 SEMINAL PLASMA PROTEIN BSP-30 KDA.

Pfam; PF00782; DSPC; 1.
SMART; SM00195; DSPC; 1.
PROSITE; PS00383; TYR PHOSPHATASE 1; 1.
PROSITE; PS0056; TYR PHOSPHATASE 2; 1.
PROSITE; PS0054; TYR PHOSPHATASE_DUAL; 1.
Hydrolase.
ACT_SITE 110 110 PHOSPHOCYCSTEINE INTERMEDIATE.
MUTAGEN 110 110 C->S: COMPLETE LOSS OF ACTIVITY.
CONFLICT 172 178 HKLKLFG -> T (IN REF. 1).
SEQUENCE 178 AA; 20627 MW; 318C99B3400A885F CRC64;

Query Match
Best Local Similarity 56.7%; Score 17; DB 1; Length 178;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXHFL 9
Db 91 YATTFL 96

RESULT 10
NU2M PARTE
ID NU2M PARTE STANDARD; PRT; 193 AA.
AC P15577;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN ND2 OR NDH2.
OS Paramesidium tetraurelia.
OG Mitochondrion.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramesidium.
OX NCBI_TaxID=5888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Stock 51;
RX MEDLINE=90174913; PubMed=2308823;
RA Pritchard A.E., Seilhamer J.J., Mahalingam R., Sable C.L.,
Venuti S.E., Cummings D.J.;
RT "Nucleotide sequence of the mitochondrial genome of Paramesidium.";
RL Nucleic Acids Res. 18:173-180 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Stock 51;
RX MEDLINE=89357489; PubMed=2670676;
RA Pritchard A.E., Venuti S.E., Ghalambor M.A., Sable C.L.,
Cummings D.J.;
RT "An unusual region of Paramesidium mitochondrial DNA containing
chloroplast-like genes.";
RL Gene 78:121-134 (1989).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
inner membrane.
CC -!- SIMILARITY: DOES NOT BELONG TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26930; AAA79255.1; -.
DR EMBL; X15917; CAA34043.1; -.
DR PIR; S07734; S07734.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 193 AA; 23181 MW; E783FD28E238AE31 CRC64;

```

```

Query Match          56.7%; Score 17; DB 1; Length 193;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      19 YSTSFL 24

Db

RESULT 11
EUTL_ECOLI
ID EUTL_ECOLI STANDARD; PRT; 219 AA.
AC P76541;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ethanolamine utilization protein eutL.
GN EUTL OR B2439.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
CC -!- FUNCTION: May be involved in the formation of a specific
CC microcompartment in the cell in which the metabolism of
CC potentially toxic by-products takes place.
CC -!- PATHWAY: Ethanolamine utilization.
CC -!- SIMILARITY: DISTANTLY RELATED TO THE CCHA/CCMK/CSOS1/PDUA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000331; AAC75492.1; -.
DR PIR; F65018;
DR EcoGene; EG14171; eutL.
DR PIRSF; PIRSF012290; EutL_PduB; 1.
KW Complete proteome.
SQ SEQUENCE 219 AA; 22788 MW; 1A5417565256243E CRC64;

Query Match          56.7%; Score 17; DB 1; Length 219;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 YXXXFL 9
      185 YSA AFL 190

Db

RESULT 12
EUTL_SALTY
ID EUTL_SALTY STANDARD; PRT; 219 AA.
AC Q9ZF09;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ethanolamine utilization protein eutL.
GN EUTL OR STM2456 OR STY2693 OR T0402.
OS Salmonella typhimurium, and

```

```

OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2;
RX MEDLINE=99395039; PubMed=10464203;
RA Kofoid E.C., Rappleye C.A., Stojiljkovic I., Roth J.R.;
RT "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium
RT encodes five homologues of carboxysome shell proteins.";
RL J. Bacteriol. 181:5317-5329 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852 (2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337 (2003).
CC -!- FUNCTION: May be involved in the formation of a specific
CC microcompartment in the cell in which the metabolism of
CC potentially toxic by-products takes place.
CC -!- PATHWAY: Ethanolamine utilization.
CC -!- SIMILARITY: DISTANTLY RELATED TO THE CCHA/CCMK/CSOS1/PDUA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF093749; AAC78125.1; -.
DR EMBL; AB008810; AAL21350.1; -.
DR EMBL; AL627274; CAD07687.1; -.
DR EMBL; AB016835; AAO68120.1; -.
DR StyGene; SG10631; eutL.
DR PIRSF; PIRSF012290; EutL_PduB; 1.
KW Complete proteome.
SQ SEQUENCE 219 AA; 22695 MW; DA50E6EDF6FCD858 CRC64;

Query Match          56.7%; Score 17; DB 1; Length 219;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;

```



Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
 DB 185 YSAFL 190

RESULT 13  
 VPR4 SALTY STANDARD; PRT; 236 AA.  
 AC P24398;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DE Virulence protein mkfA (28 kDa virulence-associated polypeptide)  
 DE (Mouse killing factor).  
 GN MKPA.  
 OS Salmonella typhimurium.  
 OC Plasmid pPI350.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS;  
 RX MEDLINE=90139869; PubMed=2559436;  
 RA Norel F., Pisano M.R., Nicoli J., Popoff M.Y.;  
 RT "Nucleotide sequence of the plasmid-borne virulence gene mkfA  
 RT encoding a 28 kDa polypeptide from Salmonella typhimurium.";  
 RL Res. Microbiol. 140:263-265(1989).  
 CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE  
 CC OF SALMONELLAS.  
 CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE  
 CC PLASMIDS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X16098; CAA34225.1; ALT\_SEQ.  
 DR PIR; A43984; A43984.  
 DR InterPro; IPR003519; Sal vir VRP3.  
 DR Pfam; PF03536; Sal vir VRP3; 1.  
 DR PRINTS; PR01342; SALVRP3PROT.  
 DR ProDom; PD017328; Sal\_vir\_VRP3; 1.  
 KW Plasmid; Virulence.  
 FT INIT MET 0 BY SIMILARITY.  
 SQ SEQUENCE 236 AA; 27006 MW; FBF344B23FFBD5C7 CRC64;

Query Match 56.7%; Score 17; DB 1; Length 236;  
 Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
 DB 167 YSASFL 172

RESULT 14  
 VPR3 SALCH STANDARD; PRT; 240 AA.  
 AC P15805;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DE 27.5 kDa virulence protein.  
 DE Salmonella cholerae-suis (Salmonella enterica).  
 OS Plasmid pKDC50.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=591;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RF-1;  
 RX MEDLINE=90192096; PubMed=2315022;  
 RA Matsui H., Kawahara K., Terakado N., Danbara H.;  
 RT "Nucleotide sequence of a gene encoding a 29 kDa polypeptide in mba  
 RT region of the virulence plasmid, pKDC50, of Salmonella  
 RT cholerae-suis.";  
 RL Nucleic Acids Res. 18:1055-1055(1990).  
 CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE  
 CC OF SALMONELLAS.  
 CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE  
 CC PLASMIDS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X51453; CAA35819.1; -.  
 DR PIR; S08402; S08402.  
 DR InterPro; IPR003519; Sal vir VRP3.  
 DR Pfam; PF03536; Sal vir VRP3; 1.  
 DR PRINTS; PR01342; SALVRP3PROT.  
 DR ProDom; PD017328; Sal\_vir\_VRP3; 1.  
 KW Plasmid; Virulence.  
 FT INIT MET 0 BY SIMILARITY.  
 SQ SEQUENCE 240 AA; 27537 MW; 5F664F56E7901D30 CRC64;

Query Match 56.7%; Score 17; DB 1; Length 240;  
 Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9  
 DB 167 YSASFL 172

RESULT 15  
 VPR3 SALTY STANDARD; PRT; 240 AA.  
 AC P21456;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 27.5 kDa virulence protein.  
 GN MKAD OR VSDD OR SPVC OR PSUT038.  
 OS Salmonella typhimurium,  
 OS Salmonella dublin, and  
 OS Salmonella enteritidis.  
 OG Plasmid pSLT, Plasmid pEX102, Plasmid pSDL2, and Plasmid pNL2001.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602, 98360, 592;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; PLASMID=pEX102;  
 RX MEDLINE=91033007; PubMed=2227425;  
 RA Taira S., Rhen M.;  
 RT "Nucleotide sequence of mkaD, a virulence-associated gene of  
 RT Salmonella typhimurium containing variable and constant regions.";  
 RL Gene 93:147-150(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RC PLASMID=PSLT;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

```
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-9.
RC SPECIES=S.typhimurium; PLASMID=pEX102;
RX MEDLINE=91244158; PubMed=2037236;
RA Taira S., Baumann M., Riikonen P., Sukupolvi S., Rhen M.;
RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
RT associated proteins of Salmonella typhimurium.";
RL FEMS Microbiol. Lett. 61:319-323(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.dublin; STRAIN=Lane; PLASMID=pSDL2;
RX MEDLINE=91251759; PubMed=2041471;
RA Krause M., Roudier C., Fierer J., Harwood J., Guiney D.;
RT "Molecular analysis of the virulence locus of the Salmonella dublin
RT plasmid pSDL2.";
RL Mol. Microbiol. 5:307-316(1991).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=AL1190; PLASMID=pNL2001;
RX MEDLINE=94362897; PubMed=8081495;
RA Suzuki S., Komase K., Matsui H., Abe A., Kawahara K., Tamura Y.,
RT Kijima M., Danbara H., Nakamura M., Sato S.;
RT "Virulence region of plasmid pNL2001 of Salmonella enteritidis.";
RL Microbiology 140:1307-1318(1994).
CC -!- FUNCTION: NOT KNOWN. THIS PROTEIN IS INVOLVED IN THE VIRULENCE
CC OF SALMONELLAS.
CC -!- SIMILARITY: TO SIMILAR PROTEIN IN OTHER SALMONELLA VIRULENCE
CC PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M34355; AAA27162.1; -.
DR EMBL; AE006471; AAL23529.1; -.
DR EMBL; X56727; CAA40050.1; -.
DR EMBL; D14490; BAA03384.1; -.
DR PIR; JQ0747; JQ0747.
DR InterPro; IPR003519; Sal_vir_VRP3.
DR Pfam; PF03536; Sal_vir_VRP3; 1.
DR PRINTS; PR01342; SALVRP3PROT.
DR ProDom; PD017328; Sal_vir_VRP3; 1.
KW Plasmid; Virulence; Complete proteome.
FT INIT MET 0
SQ SEQUENCE 240 AA; 27515 MW; 98B6C038C7395379 CRC64;

Query Match 56.7%; Score 17; DB 1; Length 240;
Best Local Similarity 50.0%; Pred.No. 1.9e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
Db 167 YSASFL 172

Search completed: October 5, 2004, 16:07:34
Job time : 12.1781 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:56:41 ; Search time 62.2466 Seconds  
(without alignments)  
81.102 Million cell updates/sec

Title: US-09-973-473a-26  
Perfect score: 30  
Sequence: 1 XXXXXXXFLXXXXXX 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	56.7	42	16 Q8FHX5	Q8fhx5 escherichia
2	17	56.7	69	2 Q9RIH9	Q9rih9 streptococc
3	17	56.7	69	16 Q8FJR6	Q8fjr6 escherichia
4	17	56.7	70	2 Q48347	Q48347 escherichia
5	17	56.7	81	17 Q8Q069	Q8q069 methanosarc
6	17	56.7	106	12 Q91ER6	Q91er6 cydia pomon
7	17	56.7	112	9 Q38579	Q38579 bacterioph
8	17	56.7	115	11 Q80UC2	Q80uc2 mus musculu
9	17	56.7	116	4 Q725K6	Q725k6 homo sapien
10	17	56.7	119	11 Q80UB6	Q80ub6 mus musculu
11	17	56.7	126	5 Q7YV69	Q7yv69 trypanosoma
12	17	56.7	141	16 Q92BY4	Q92by4 xylella fas
13	17	56.7	141	16 Q87D78	Q87d78 xylella fas
14	17	56.7	144	16 Q8RDZ6	Q8rdz6 fusobacteri
15	17	56.7	148	7 Q9XRE7	Q9xre7 ovis aries
16	17	56.7	153	16 Q9CJX7	Q9cjx7 pasteurella

17	56.7	154	10	Q8H3P6	Q8h3p6 oryza sativ
18	17	56.7	156	16 Q9CCP2	Q9ccp2 mycobacteri
19	17	56.7	159	16 Q8FU49	Q8fu49 corynebacte
20	17	56.7	166	4 Q8NAC1	Q8nac1 homo sapien
21	17	56.7	172	2 Q68298	Q68298 borrelia af
22	17	56.7	172	5 Q01458	Q01458 caenorhabdi
23	17	56.7	173	12 Q85315	Q85315 shope fibro
24	17	56.7	174	2 Q8GCB7	Q8gcb7 borrelia af
25	17	56.7	194	16 Q8XWP9	Q8xwp9 ralstonia s
26	17	56.7	196	16 Q81JK6	Q81jk6 bacillus an
27	17	56.7	196	16 Q814I5	Q814i5 bacillus ce
28	17	56.7	199	10 Q84NQ7	Q84nq7 oryza sativ
29	17	56.7	216	16 Q89N63	Q89n63 bradyrhizob
30	17	56.7	219	16 Q8XB11	Q8xb11 escherichia
31	17	56.7	219	16 Q8FFA2	Q8ffa2 escherichia
32	17	56.7	219	16 Q83K70	Q83k70 shigella fl
33	17	56.7	223	5 Q02255	Q02255 caenorhabdi
34	17	56.7	227	5 Q9W3Z6	Q9w3z6 drosophila
35	17	56.7	228	17 Q29279	Q29279 archaeoglob
36	17	56.7	232	2 Q83ZD9	Q83zd9 gamma-prote
37	17	56.7	241	2 Q04251	Q04251 salmonella
38	17	56.7	245	16 Q98QY3	Q98qy3 mycoplasma
39	17	56.7	246	10 Q9M234	Q9m234 arabidopsis
40	17	56.7	250	16 Q8Y3U9	Q8y3u9 listeria mo
41	17	56.7	253	16 Q8R803	Q8r803 thermoaer
42	17	56.7	260	5 Q81DZ4	Q81dz4 plasmodium
43	17	56.7	262	16 Q8EN08	Q8en08 oceanobacil
44	17	56.7	273	5 Q815I0	Q815i0 plasmodium
45	17	56.7	275	17 Q97CD5	Q97cd5 thermoplas

#### ALIGNMENTS

#### RESULT 1

Q8FHXS PRELIMINARY; PRT; 42 AA.  
AC Q8FHXS; TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN Cl702.  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
RX MEDLINE=22386234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
DR EMBL: AE016760; AAN80169.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 42 AA; 4739 MW; F0571A4C8EE18B01 CRC64;

Query Match 56.7%; Score 17; DB 16; Length 42;  
Best Local Similarity 50.0%; Pred. No. 4.2e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXAPL 9  
Db 20 YAAAPL 25

#### RESULT 2

Q9RIH9

```
ID Q9RIH9 PRELIMINARY; PRT; 69 AA.
AC Q9RIH9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS101;
RA Podbielski A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U78969; AAF08327.1; -.
KW Hypothetical protein.
SQ SEQUENCE 69 AA; 7189 MW; E717113F2C3558B9 CRC64;

Query Match 56.7%; Score 17; DB 2; Length 69;
Best Local Similarity 50.0%; Pred. NO. 6.5e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
DB 14 YASFL 19

RESULT 3
Q8FJR6 PRELIMINARY; PRT; 69 AA.
ID Q8FJR6 PRELIMINARY; PRT; 69 AA.
AC Q8FJR6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN C0839.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22398234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016757; AAN79312.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 69 AA; 7938 MW; A26139712997DA90 CRC64;

Query Match 56.7%; Score 17; DB 16; Length 69;
Best Local Similarity 50.0%; Pred. NO. 6.5e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
DB 4 YTSFL 9

RESULT 4
Q48347 PRELIMINARY; PRT; 70 AA.
ID Q48347 PRELIMINARY; PRT; 70 AA.
AC Q48347;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Coli transposase element variant IS903.B present on kanamycin
```

```
DE resistance transposon Tn 2680.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85295477; PubMed=2993802;
RA Mollet B., Iida S., Arber W.;
RT "An active variant of the prokaryotic transposable element IS903
carries an amber stop codon in the middle of an open reading frame.";
RL Mol. Gen. Genet. 199:534-536(1985).
DR EMBL; X02527; CAA26363.1; -.
DR PIR; I77547; I77547.
SQ SEQUENCE 70 AA; 7633 MW; D01541A4032FD53F CRC64;

Query Match 56.7%; Score 17; DB 2; Length 70;
Best Local Similarity 50.0%; Pred. NO. 6.5e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
DB 14 YTAFL 19

RESULT 5
Q8Q069 PRELIMINARY; PRT; 81 AA.
ID Q8Q069 PRELIMINARY; PRT; 81 AA.
AC Q8Q069;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Conserved protein.
GN MM0268.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wietzer A., Baumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
DR EMBL; AE013251; AAM29964.1; -.
KW Complete proteome.
SQ SEQUENCE 81 AA; 9140 MW; BEE0AEA3A1E63C54 CRC64;

Query Match 56.7%; Score 17; DB 17; Length 81;
Best Local Similarity 50.0%; Pred. NO. 7.4e+02;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXXFL 9
DB 60 YSSAFL 65

RESULT 6
Q91ER6 PRELIMINARY; PRT; 106 AA.
ID Q91ER6 PRELIMINARY; PRT; 106 AA.
AC Q91ER6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF139.
GN ORF139.
OS Cydia pomonella granulosis virus (CpGV) (Cydia pomonella
```

OS granulovirus).  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.  
 OX NCBI\_TaxID=28289;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mexican 1;  
 RX MEDLINE=93188168; PubMed=8445726;  
 RA Crook N.E., Clem R.J., Miller L.K.;  
 RT "An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif.";  
 RL J. Virol. 67:2168-2174(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mexican 1;  
 RX MEDLINE=96207404; PubMed=8615018;  
 RA Theilmann D.A., Chantler J.K., Stewart S., Flipse H.T., Vlak J.M.,  
 RA Crook N.E.;  
 RT "Characterization of a highly conserved baculovirus structural protein that is specific for occlusion-derived virions.";  
 RL Virology 218:148-158(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mexican 1;  
 RX MEDLINE=97380577; PubMed=9237352;  
 RA Kang W., Crook N.E., Winstanley D., O'Reilly D.R.;  
 RT "Complete sequence and transposon mutagenesis of the BamHI J fragment of Cydia pomonella granulosis virus.";  
 RL Virus Genes 14:131-136(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mexican 1;  
 RX MEDLINE=98418511; PubMed=9747739;  
 RA Kang W., Tristem M., Maeda S., Crook N.E., O'Reilly D.R.;  
 RT "Identification and characterization of the Cydia pomonella granulovirus cathepsin and chitinase genes.";  
 RL J. Gen. Virol. 79:2283-2292(1998).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mexican 1;  
 RA Luque T., Finch R., Crook N., O'Reilly D.R., Winstanley D.;  
 RT "The complete sequence of the Cydia pomonella granulovirus genome.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U53466; AAK70799.1; -;  
 SQ SEQUENCE 106 AA; 11671 MW; 47B4E1987B97848E CRC64;  
 Query Match 56.7%; Score 17; DB 12; Length 106;  
 Best Local Similarity 50.0%; Pred. No. 9.4e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 YXXSFL 9  
 Db 47 YTSFL 52  
 RESULT 7  
 Q38579  
 ID Q38579 PRELIMINARY; PRT; 112 AA.  
 AC Q38579;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Product required for head morphogenesis.  
 GN 9.  
 OS Bacteriophage SPPI.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
 OC Lambda-like viruses.  
 OX NCBI\_TaxID=10724;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Becker B., Gassel M., Tavares P., Lurz R., Alonso J.C.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Alonso J.C., Luder G., Stiege A.C., Chai S., Weise F., Trautner T.A.;  
 RN Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RA Alonso J.C.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X89721; CAA61867.1; -;  
 DR EMBL; X97918; CAA66587.1; -;  
 DR PIR; S58139; S58139.  
 SQ SEQUENCE 112 AA; 12633 MW; 143BFCB736B2D52 CRC64;  
 Query Match 56.7%; Score 17; DB 9; Length 112;  
 Best Local Similarity 50.0%; Pred. No. 9.9e+02;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 YXXSFL 9  
 Db 105 YTSFL 110  
 RESULT 8  
 Q80UC2  
 ID Q80UC2 PRELIMINARY; PRT; 115 AA.  
 AC Q80UC2;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Metabotropic glutamate receptor 4 (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22584407; PubMed=12679517;  
 RA Vassilatis D.K., Hohmann J.G., Zeng H., Li F., Ranchalis J.E.,  
 RA Mortrud M.T., Brown A., Rodriguez S.S., Weller J.R., Wright A.C.,  
 RA Bergmann J.E., Gaitanaris G.A.;  
 RT "The G Protein-Coupled Receptor Repertoires of Human and Mouse.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:4903-4908(2003).  
 DR EMBL; AY255558; AA085070.1; -;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008067; F:metabotropic glutamate, GABA-B-like recepto. .; IEA.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR000337; GPCR\_Mgr.  
 DR PROSITE; PS00980; G\_PROTEIN\_RECEP\_F3\_2; 1.  
 DR PROSITE; PS50259; G\_PROTEIN\_RECEP\_F3\_4; 1.  
 KW Receptor.  
 FT NON\_TER 1 1  
 FT NON\_TER 115 115  
 SQ SEQUENCE 115 AA; 13082 MW; 3487954F5AB9C296 CRC64;  
 Query Match 56.7%; Score 17; DB 11; Length 115;  
 Best Local Similarity 50.0%; Pred. No. 1e+03;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 YXXSFL 9  
 Db 110 YATFL 115  
 RESULT 9  
 Q725K6  
 ID Q725K6 PRELIMINARY; PRT; 116 AA.  
 AC Q725K6;  
 DT 01-OCT-2003 (TRENBLrel. 25, Created)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE ARM-repeat protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;

```

RN  SEQUENCE FROM N.A.
RP  TISSUE=Testis;
RA  Xu Z.Y., Xu M., Yin L.L., Lu L., Li J.M., Zhou J.M., Sha J.H.;
RT  "An ARM-repeat superfamily protein acts in spermatogenesis.";
RL  Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AY313779; AAP81010.1; -.
SQ  SEQUENCE 116 AA; 13666 MW; 67C4FCA7023042AF CRC64;

Query Match          56.7%; Score 17; DB 4; Length 116;
Best Local Similarity 50.0%; Pred. No. 1e+03; 3; Indels 0; Gaps 0;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  4 YXXXFL 9
Db  36 YSTSFL 41

RESULT 10
Q80UD6 PRELIMINARY; PRT; 119 AA.
AC  Q80UD6;
DT  01-JUN-2003 (TrEMBLrel. 24, Created)
DT  01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  G protein-coupled receptor SALPR (Fragment).
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  SEQUENCE FROM N.A.
RP  MEDLINE=22584407; PubMed=12679517;
RA  Vassiliadis D.K., Hohmann J.G., Zeng H., Li F., Ranchalis J.E.,
RA  Mortrud M.T., Brown A., Rodriguez S.S., Weller J.R., Wright A.C.,
RA  Bergmann J.E., Gaitanaris G.A.;
RT  "The G Protein-Coupled Receptor Repertoires of Human and Mouse.";
RL  Proc. Natl. Acad. Sci. U.S.A. 100:4903-4908 (2003).
DR  EMBL; AY255339; AAO80501.1; -.
DR  GO; GO:0016021; C:integral to membrane; IEA.
DR  GO; GO:0004872; F:receptor activity; IEA.
DR  GO; GO:0003584; F:rhodopsin-like receptor activity; IEA.
DR  GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR  InterPro; IPR000276; P:G-protein coupled receptor protein signalin. .; IEA.
DR  PRINTS; PR00237; GPCRHHODPSN.
DR  PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW  Receptor.
FT  NON_TER 1
FT  NON_TER 119
SQ  SEQUENCE 119 AA; 12632 MW; DF1D9BB389245D73 CRC64;

Query Match          56.7%; Score 17; DB 11; Length 119;
Best Local Similarity 50.0%; Pred. No. 1e+03; 3; Indels 0; Gaps 0;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  4 YXXXFL 9
Db  86 YASTFL 91

RESULT 11
Q7YV69 PRELIMINARY; PRT; 126 AA.
AC  Q7YV69;
DT  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Hypothetical protein.
OS  Trypanosoma brucei.
GN  TB927.2.3630.
OC  Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX  NCBI_TaxID=5691;
RN  [1]

```

```

RP  SEQUENCE FROM N.A.
RC  STRAIN=GUTat10.1;
RA  El-Sayed N.M.A., Ghedin E., Song J., Macleod A., Bringaud P.,
RA  Larkin C., Wanless D., Peterson J., Hou L., Taylor S., Tweedie A.,
RA  Biteau N., Khalak H.G., Lin X., Mason T., Hannick L., Caler E.,
RA  Blandin G., Bartholomeu D., Simpson A.J., Kaul S., Zhao H., Pai G.,
RA  Van Aken S., Utterback T., Haas B., Koo H.L., Umayam L., Suh B.,
RA  Gerard C., Leech V., Qi R., Zhou S., Schwartz D., Feldblum T.,
RA  Salzberg S., Tait A., Turner M.R., Ullu E., White O., Melville S.,
RA  Adams M.D., Fraser C.M., Donelson J.E.;
RT  "The sequence and analysis of Trypanosoma brucei chromosome II.";
RL  Nucleic Acids Res. 0:0-0(2003).
DR  EMBL; AE017169; RAQ15820.1; -.
KW  Hypothetical protein.
SQ  SEQUENCE 126 AA; 14812 MW; F8F1A4EB0102B2DD CRC64;

Query Match          56.7%; Score 17; DB 5; Length 126;
Best Local Similarity 50.0%; Pred. No. 1.1e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  4 YXXXFL 9
Db  16 YATTF 21

RESULT 12
Q9PBV4 PRELIMINARY; PRT; 141 AA.
AC  Q9PBV4;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Hypothetical protein XF2001.
GN  Xylella fastidiosa.
OS  Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC  Xanthomonadaceae; Xylella.
OX  NCBI_TaxID=2371;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=9a5c;
RX  MEDLINE=20365717; PubMed=10910347;
RA  Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA  Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA  Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA  Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA  Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
RA  Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA  Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA  Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA  Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA  Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper B.L., Kitajima J.P.,
RA  Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,
RA  Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA  Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA  Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA  Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA  Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA  Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA  de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA  Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA  Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA  de Rosa V.B. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA  da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA  da Silva J.P., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA  de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA  Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA  Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT  "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL  Nature 406:151-159 (2000).
DR  EMBL; AE004019; AAF84803.1; -.
DR  FIR; C82612; C82612.
KW  Hypothetical protein; Complete proteome.

```

```
SQ SEQUENCE 141 AA; 15298 MW; 01D9D650F566548 CRC64;
Query Match 56.7%; Score 17; DB 16; Length 141;
Best Local Similarity 50.0%; Pred. NO. 1.2e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9
DB 128 YAAFL 133

RESULT 13
Q87D78 PRELIMINARY; PRT; 141 AA.
AC Q87D78;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein.
GN PD0808.
OS Xylella fastidiosa (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.B.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Bianco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.P., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa.";
RL J. Bacteriol. 185:1018-1026 (2003).
DR EMBL; AE012556; AA028676.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 141 AA; 15279 MW; EE23650F5666553 CRC64;

Query Match 56.7%; Score 17; DB 16; Length 141;
Best Local Similarity 50.0%; Pred. NO. 1.2e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9
DB 128 YAAFL 133

RESULT 14
Q8RD26 PRELIMINARY; PRT; 144 AA.
AC Q8RD26;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein FN1344.
GN FN1344.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
```

```
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018 (2002).
DR EMBL; AE010639; AAL95540.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 144 AA; 17114 MW; 3711CAF5B7900F4B CRC64;

Query Match 56.7%; Score 17; DB 16; Length 144;
Best Local Similarity 50.0%; Pred. NO. 1.2e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 YXXFL 9
DB 8 YSSTFL 13

RESULT 15
Q9XRE7 PRELIMINARY; PRT; 148 AA.
AC Q9XRE7;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE MHC class II antigen (Fragment).
GN OLA-DQ42.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RX SEQUENCE FROM N.A.
RP STRAIN=Corriedale;
RC MEDLINE=97083726; PubMed=8930070;
RA Escayg A.P., Hickford J.G., Montgomery G.W., Dodds K.G., Bullock D.W.;
RT "Polymorphism at the ovine major histocompatibility complex class II
RT loci.";
RL Anim. Genet. 27:305-312 (1996).
RN [2]
RX SEQUENCE FROM N.A.
RP STRAIN=Corriedale;
RC Slow S., Ridgway H.J., Hickford J.G.H.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF129119; AAD33405.1; -.
DR HSSP; P01910; 1IAK.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0045012; F:MHC class II receptor activity; IEA.
DR GO; GO:0019884; P:antigen presentation, exogenous antigen; IEA.
DR GO; GO:0019886; P:antigen processing, exogenous antigen via M. .; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001003; MHC_II_alpha.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00407; IGG1_1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 148
SQ SEQUENCE 148 AA; 16974 MW; A9AB27B0A10C9CF2 CRC64;

Query Match 56.7%; Score 17; DB 7; Length 148;
Best Local Similarity 50.0%; Pred. NO. 1.3e+03;
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 4 YXXFL 9  
| | |  
Db 113 YATSFL 118

Search completed: October 5, 2004, 16:12:35  
Job time : 65.2466 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:06 ; Search time 35.9589 Seconds  
(without alignments)  
55.003 Million cell updates/sec

Title: US-09-973-473A-28

Perfect score: 18

Sequence: 1 YXXFPXX 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	72.2	6	ADA89445	Ada89445 Hypersens
2	13	72.2	7	AAE31849	AAE31849 Androgen
3	13	72.2	9	AAR89362	AAR89362 Immunogen
4	13	72.2	9	AAV05025	AAV05025 Tumour an
5	13	72.2	9	AAV98477	AAV98477 HLA class
6	13	72.2	9	AAV98475	AAV98475 HLA class
7	13	72.2	9	AAV98476	AAV98476 HLA class
8	13	72.2	9	AAV98474	AAV98474 HLA class
9	13	72.2	9	AAV984528	AAV984528 Human leu
10	13	72.2	9	AAV06288	AAV06288 Human leu
11	13	72.2	9	AAV68280	AAV68280 Human leu
12	13	72.2	9	ABP25096	ABP25096 Human MHC
13	13	72.2	9	ABP25090	ABP25090 Human MHC
14	13	72.2	9	ABP25092	ABP25092 Human MHC
15	13	72.2	9	ABP11509	ABP11509 HLA class
16	13	72.2	9	ABP25094	ABP25094 Human MHC
17	13	72.2	9	ABP25091	ABP25091 Human MHC
18	13	72.2	9	ABP25095	ABP25095 Human MHC
19	13	72.2	9	AAV99376	AAV99376 Human leu
20	13	72.2	9	AAV03825	AAV03825 Hepatitis
21	13	72.2	9	AAJ00066	AAJ00066 Hepatitis
22	13	72.2	9	AAJ03826	AAJ03826 Hepatitis
23	13	72.2	9	AAJ03820	AAJ03820 Hepatitis
24	13	72.2	9	AAJ03824	AAJ03824 Hepatitis
25	13	72.2	9	AAJ00068	AAJ00068 Hepatitis

26	13	72.2	9	AAJ03821	AAJ03821 Hepatitis
27	13	72.2	9	AAJ00067	AAJ00067 Hepatitis
28	13	72.2	9	AAJ00069	AAJ00069 Hepatitis
29	13	72.2	9	AAJ03822	AAJ03822 Hepatitis
30	13	72.2	9	AAJ09694	AAJ09694 Hepatitis
31	13	72.2	9	ABJ09689	ABJ09689 Hepatitis
32	13	72.2	9	ABJ09692	ABJ09692 Hepatitis
33	13	72.2	9	ABJ09690	ABJ09690 Hepatitis
34	13	72.2	9	ABJ05800	ABJ05800 Hepatitis
35	13	72.2	9	ABJ05798	ABJ05798 Hepatitis
36	13	72.2	9	ABJ09693	ABJ09693 Hepatitis
37	13	72.2	9	ABJ09690	ABJ09690 Hepatitis
38	13	72.2	9	ABJ09688	ABJ09688 Hepatitis
39	13	72.2	9	ABJ05801	ABJ05801 Hepatitis
40	13	72.2	9	ABJ05799	ABJ05799 Hepatitis
41	13	72.2	9	ABJ37976	ABJ37976 Human cyt
42	13	72.2	12	ADC82768	ADC82768 CDR regio
43	13	72.2	12	AAW80391	AAW80391 Peptide e
44	13	72.2	12	AAV58436	AAV58436 Staphyloc
45	13	72.2	13	AAW58716	AAW58716 Tryptic 4

## ALIGNMENTS

### RESULT 1

ADA89445

ID ADA89445 standard; peptide; 6 AA.

XX AC ADA89445;

XX DT 20-NOV-2003 (first entry)

XX DE Hypersensitive response elicitor receptor AhrBp1p peptide SEQ ID NO:70.

XX KW receptor; plant; plant pathogen hypersensitive response elicitor;

XX KW hypersensitive response elicitor; disease resistance;

XX KW enhancing plant growth; controlling insect; stress tolerance.

XX OS Arabidopsis thaliana.

XX PN WO2003054211-A2.

XX PD 03-JUL-2003.

XX PF 31-OCT-2002; 2002WO-US035252.

XX PR 31-OCT-2001; 2001US-0335776P.

XX PR 17-JUN-2002; 2002US-00174209.

XX (EDEN-) EDEN BIOSCIENCE CORP.

XX Song X, Bariola PA, Linderth NA, Fan H, Wei Z;

XX WPI; 2003-559153/52.

XX New isolated protein as a receptor in plants for plant pathogen

XX PT hypersensitive response elicitors, useful for identifying agents that

XX PT impart disease resistance, enhance plant growth, control insects and/or

XX PT impart stress tolerance.

XX Claim 12; Page 56; 104pp; English.

XX The present invention describes an isolated protein (1) serving as a

XX receptor in plants for plant pathogen hypersensitive response elicitors.

XX Also described: (1) an isolated nucleic acid encoding (1); (2) an

XX antisense nucleic acid molecule to the nucleic acid of (1); (3) an

XX expression vector containing the nucleic acid of (1) heterologous to the

XX expression vector; (4) a transgenic host cell transformed with the

XX nucleic acid of (1) or the DNA molecule of (2); (5) a transgenic plant

XX transformed with the nucleic acid of (1) or the DNA molecule of (2); (6)

XX identifying agents targeting plant cells, comprising forming a reaction

XX mixture having (1) or the host cell of (4) and a candidate agent,

CC evaluating the reaction mixture for binding between (I) or the protein  
 CC produced by the host cell and the candidate agent, and identifying  
 CC candidate compounds which bind to the proteins in the reaction mixture as  
 CC plant cell targeting agents; (7) enhancing plant receptivity to treatment  
 CC with hypersensitive response elicitors, comprising providing a transgenic  
 CC plant or plant seed transformed with the nucleic acid of (1); and (8)  
 CC imparting disease resistance, enhancing growth, controlling insects,  
 CC and/or imparting stress resistance to plants, comprising providing a  
 CC transgenic plant or plant seed transformed with a DNA construct effective  
 CC to silence expression of a nucleic acid molecule of (1), or transformed  
 CC with the nucleic acid molecule of (1). The methods and compositions of  
 CC the invention are useful for identifying agents targeting plant cells to  
 CC enhance a plant's receptivity to treatment with a hypersensitive response  
 CC elicitor. The hypersensitive response elicitor treatment includes  
 CC imparting disease resistance, enhancing plant growth, controlling insects  
 CC and/or imparting stress tolerance. The present sequence represents an  
 CC Arabidopsis thaliana hypersensitive response elicitor receptor AtHRP1p  
 CC peptide, which is used in an example from the present invention.  
 XX  
 SQ Sequence 6 AA;

Query Match 72.2%; Score 13; DB 6; Length 6;  
 Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXKF 5  
 |  
 Db 1 YSSAF 5

RESULT 2  
 AAE31849  
 ID AAE31849 standard; peptide; 7 AA.

XX  
 AC AAE31849;

XX  
 DT 07-MAR-2003 (first entry)

XX  
 DE Androgen receptor binding peptide #100.

XX Androgen receptor; androgen-associated disorder; prostate cancer; acne;  
 KW benign prostatic hypertrophy; hirsutism; androgen insensitivity syndrome;  
 KW male pattern baldness; Stein-Leventhal syndrome; infertility; cytostatic;  
 KW X-linked spinal bulbar muscular atrophy; antiseborrheic; dermatological;  
 KW depilatory; androgen receptor binding peptide.

XX  
 OS Unidentified.

XX  
 FH Location/Qualifiers

FT Misc-difference 1 /note= "Linked to Xa-Y1; Where Xa is independently a  
 FT direct bond or a peptidic structure comprising from about  
 FT 1-25 amino acid residues and Y1 is hydrogen, alkyl or  
 FT acyl"

FT Misc-difference 7 /note= "Linked to Xb-Y2; Where Xb is independently a  
 FT direct bond or a peptidic structure comprising from about  
 FT 1-25 amino acid residues and Y2 is -OH, amino or  
 FT monosubstituted or disubstituted amino"

XX  
 PN WO200272612-A2.

XX  
 PD 19-SEP-2002.

XX  
 PF 12-MAR-2002; 2002WO-US007487.

XX  
 PR 12-MAR-2001; 2001US-0275240P.

XX  
 PR 28-JAN-2002; 2002US-0352399P.

XX  
 PA (PRAE-) PRAECIS PHARM INC.

XX  
 PI Joyal JL, Mueller J, Oza VB, Findeis MA;

XX

DR WPI; 2003-067363/06.  
 XX  
 PT New peptide modulators of androgen receptor, useful for treating androgen  
 PT -associated disorder, e.g. prostate cancer, particularly hormonally  
 PT refractive prostate cancer, colon cancer, lung cancer, acne, or  
 PT hirsutism.

XX  
 PS Claim 29; Page 36; 68pp; English.

XX  
 CC The present invention relates to novel peptide modulators of androgen  
 CC receptor. The peptides of the invention are useful for treating androgen-  
 CC associated disorders such as prostate cancer, particularly hormonally  
 CC refractive prostate cancer, colon cancer, lung cancer, benign prostatic  
 CC hypertrophy, acne, hirsutism, male pattern baldness, Stein-Leventhal  
 CC syndrome, androgen insensitivity syndrome, infertility, endometrial  
 CC cancer and X-linked spinal bulbar muscular atrophy. The present sequence  
 CC is an androgen receptor binding peptide

XX  
 SQ Sequence 7 AA;

Query Match 72.2%; Score 13; DB 6; Length 7;  
 Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXKF 5  
 |  
 Db 1 YAAAF 5

RESULT 3

AAR89362  
 ID AAR89362 standard; peptide; 9 AA.

XX  
 AC AAR89362;

XX  
 DT 18-SEP-1996 (first entry)

XX  
 DE Immunogenic peptide, based on B35 consensus peptide.

XX Immunogenic peptide; supermotif; HLA molecule; CTL response; therapeutic;  
 KW diagnostic; cancer; viral infection; hepatitis B; hepatitis C.

XX  
 OS Synthetic.

XX  
 PN WO9603140-A1.

XX  
 PD 08-FEB-1996.

XX  
 PF 21-JUL-1995; 95WO-US009234.

XX  
 PR 21-JUL-1994; 94US-00278634.

XX  
 PR 23-NOV-1994; 94US-00344824.

XX  
 PR 30-MAY-1995; 95US-00452843.

XX  
 PA (CYTE-) CYTEL CORP.

XX  
 PI Sette A, Sidney J;

XX  
 WPI; 1996-116784/12.

XX  
 FT Compen. comprising immunogenic peptide with supermotif allowing more than  
 FT one HLA mol. to bind - used to induce CTL response in patient and for in  
 FT vivo and ex vivo therapeutic and diagnostic applications.

XX  
 PS Claim 2; Page 26; 32pp; English.

XX  
 CC The sequences given in AAR89362-82 are immunogenic peptides which were  
 CC use in the composition of the invention. The composition comprises an  
 CC immunogenic peptide of 9-10 residues with a supermotif which allows  
 CC binding of more than one HLA molecule. It pref. comprises two conserved  
 CC residues, a first at the 2nd position from the N- terminal is Pro, and a  
 CC 2nd at the C-terminal is Met. These peptides are used to induce a CTL  
 CC response in a patient. They are also useful in compositions for in vivo

CC and ex vivo therapeutic and diagnostic applications, e.g the treatment of  
CC cancer and viral infections, e.g. hepatitis B and C  
XX  
SQ Sequence 9 AA;

Query Match 72.2%; Score 13; DB 2; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
|  
5 YAAAF 9

Db

RESULT 4  
AAAY05025  
ID AAY05025 standard; peptide; 9 AA.

XX  
AC AAY05025;  
XX  
DT 16-JUN-1999 (first entry)

XX Tumour antigen antibody light chain CDR3 clone F3.

DE  
XX Tumour antigen; antibody; CDR; complementarity determining region;  
KW binding molecule identification; tumour-specific binding polypeptide;  
KW cancer therapy; light chain.

XX Homo sapiens.

OS

XX WO9906834-A2.

PN

XX 11-FEB-1999.

PD

XX 04-AUG-1998; 98WO-US016280.

PF

XX 04-AUG-1997; 97US-00905825.

PR

XX (IXSY-) IXSYS INC.

PA

XX Watkins JD, Huse WD, Wu H;

PI

XX WPI; 1999-153951/13.

DR

XX N-PSDB; AAX28202.

DR

XX Identifying binding molecules for ligands, particularly tumour antigens -  
PT by selectively immobilising a population of binding molecules to a solid  
PT support and screening for binding to two or more ligands.

XX Claim 15; Page 57; 80pp; English.

PS

XX This sequence represents a light chain complementarity determining region  
CC (CDR) from a tumour antigen specific antibody. The invention relates to a  
CC method for identifying a binding molecule having selective affinity for a  
CC ligand comprising: (a) selectively immobilising a diverse population of  
CC binding molecules to a solid support; (b) simultaneously contacting the  
CC diverse population immobilised on the solid support with 2 or more  
CC ligands; and (c) determining at least one binding molecule which  
CC selectively binds to one or more of the ligands. The method allows for  
CC the rapid and efficient methods for the identification of binding  
CC molecules which exhibit selective affinity for one or more ligands of  
CC interest. They are used particularly for identifying tumour-specific  
CC binding polypeptides which can be used as targeting agents for cancer  
CC therapy that minimises impact on non-tumour tissues

XX Sequence 9 AA;

Query Match 72.2%; Score 13; DB 2; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
|

Db

Db 3 YSSTF 7

RESULT 5

AAAB98477

ID AAB98477 standard; protein; 9 AA.

XX

AC AAB98477;

XX

DT 22-AUG-2001 (first entry)

XX

DE HLA class I standard peptide binding affinity B\*5401.

XX

KW Human papillomavirus; human leukocyte antigen; HLA; immune response; HPV;  
KW epitope; T cell; identification; vaccine; infection; genital wart;  
KW neoplastic growth; antiviral.

XX

OS Homo sapiens.

XX

PN WO200141799-A1.

XX

PD 14-JUN-2001.

XX

PF 11-DEC-2000; 2000WO-US033549.

XX

PR 10-DEC-1999; 99US-0172705P.

PR

XX 15-AUG-2000; 2000US-00641528.

XX

PA (EPIM-) EPIMUNE INC.

XX

PI Sette A, Sidney J, Southwood S, Chesnut R, Celis E, Grey HM;

XX

DR WPI; 2001-381497/40.

XX

PT An isolated human papilloma virus (HPV) epitope, useful in vaccines for  
PT treating HPV infections.

XX

PS Disclosure; Page 98; 756pp; English.

XX

CC The present invention describes an isolated prepared human papillomavirus  
CC (HPV) epitope (I). (I) has antiviral activity, and can be used in vaccine  
CC production. Peptides and corresponding nucleic acid compositions from the  
CC present invention are useful for stimulating an immune response to HPV by  
CC stimulating the production of CTL or HTL responses, specifically in the  
CC treatment or prophylaxis of HPV infection, in persons who have not  
CC manifested symptoms e.g. genital warts or neoplastic growth. The peptides  
CC can also be used in a tetramer staining assay to assess peripheral blood  
CC mononuclear cells for the presence of antigen-specific CTLs following  
CC exposure to a pathogen or immunogen, and as reagents to evaluate immune  
CC recall responses or evaluate the efficacy of a vaccine. The vaccine  
CC compositions are useful for removing warts or treating HPV infections.  
CC The epitopes for inclusion in an epitope-base vaccine may be selected  
CC from conserved regions of viral or tumour-associated antigens, which  
CC reduces the likelihood of escape mutants, also immunosuppressive epitopes  
CC that may be present in whole antigens can be avoided with the use of  
CC epitope-base vaccines. An additional advantage is the ability to combine  
CC selected epitopes (CTL and HTL) and to modify the composition of the  
CC epitopes achieving enhanced immunogenicity, the major benefit of the  
CC vaccine is that is safe and efficacious. AAB98477 to AAB98477 represent  
CC polypeptide sequences used in the exemplification of the present  
CC invention

XX Sequence 9 AA;

Query Match 72.2%; Score 13; DB 4; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
|

Db 5 YAAAF 9

```

RESULT 6
AAB98475
ID AAB98475 standard; protein; 9 AA.
XX
AC AAB98475;
XX
AC AAB98475;
XX
DT 22-AUG-2001 (first entry)
XX
DE HLA class I standard peptide binding affinity B51.
XX
KW Human papillomavirus; human leukocyte antigen; HLA; immune response; HPV;
KW epitope; T cell; identification; vaccine; infection; genital wart;
KW neoplastic growth; antiviral.
XX
OS Homo sapiens.
XX
PN WO200141799-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033549.
XX
PR 10-DEC-1999; 99US-0172705P.
XX
PR 15-AUG-2000; 2000US-00641528.
XX
PA (EPIN-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Chesnut R, Celis E, Grey HM;
XX
PI WPI; 2001-381497/40.
XX
DR An isolated human papilloma virus (HPV) epitope, useful in vaccines for
XX PT treating HPV infections.
XX
PS Disclosure; Page 98; 756pp; English.
XX
CC The present invention describes an isolated prepared human papillomavirus
XX (HPV) epitope (I). (I) has antiviral activity, and can be used in vaccine
XX production. Peptides and corresponding nucleic acid compositions from the
XX present invention are useful for stimulating an immune response to HPV by
XX stimulating the production of CTL or HTL responses, specifically in the
XX treatment or prophylaxis of HPV infection, in persons who have not
XX manifested symptoms e.g. genital warts or neoplastic growth. The peptides
XX can also be used in a tetramer staining assay to assess peripheral blood
XX mononuclear cells for the presence of antigen-specific CTLs following
XX exposure to a pathogen or immunogen, and as reagents to evaluate immune
XX recall responses or evaluate the efficacy of a vaccine. The vaccine
XX compositions are useful for removing warts or treating HPV infections.
XX The epitopes for inclusion in an epitope-base vaccine may be selected
XX from conserved regions of viral or tumour-associated antigens, which
XX reduces the likelihood of escape mutants, also immunosuppressive epitopes
XX that may be present in whole antigens can be avoided with the use of
XX epitope-base vaccines. An additional advantage is the ability to combine
XX selected epitopes (CTL and HTL) and to modify the composition of the
XX epitopes achieving enhanced immunogenicity, the major benefit of the
XX vaccine is that is safe and efficacious. AAB98391 to AAB98477 represent
XX polypeptide sequences used in the exemplification of the present
XX invention
XX
SQ Sequence 9 AA;
Query Match 72.2%; Score 13; DB 4; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.3e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 YXXXF 5
Db 5 YAAAF 9
RESULT 7
AAB98476
ID AAB98476 standard; protein; 9 AA.
XX
AC AAB98476;
XX
DT 22-AUG-2001 (first entry)
XX
DE HLA class I standard peptide binding affinity B*5301.
XX
KW Human papillomavirus; human leukocyte antigen; HLA; immune response; HPV;
KW epitope; T cell; identification; vaccine; infection; genital wart;
KW neoplastic growth; antiviral.
XX
OS Homo sapiens.
XX
PN WO200141799-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US033549.
XX
PR 10-DEC-1999; 99US-0172705P.
XX
PR 15-AUG-2000; 2000US-00641528.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S, Chesnut R, Celis E, Grey HM;
XX
PI WPI; 2001-381497/40.
XX
DR An isolated human papilloma virus (HPV) epitope, useful in vaccines for
XX PT treating HPV infections.
XX
PS Disclosure; Page 98; 756pp; English.
XX
CC The present invention describes an isolated prepared human papillomavirus
XX (HPV) epitope (I). (I) has antiviral activity, and can be used in vaccine
XX production. Peptides and corresponding nucleic acid compositions from the
XX present invention are useful for stimulating an immune response to HPV by
XX stimulating the production of CTL or HTL responses, specifically in the
XX treatment or prophylaxis of HPV infection, in persons who have not
XX manifested symptoms e.g. genital warts or neoplastic growth. The peptides
XX can also be used in a tetramer staining assay to assess peripheral blood
XX mononuclear cells for the presence of antigen-specific CTLs following
XX exposure to a pathogen or immunogen, and as reagents to evaluate immune
XX recall responses or evaluate the efficacy of a vaccine. The vaccine
XX compositions are useful for removing warts or treating HPV infections.
XX The epitopes for inclusion in an epitope-base vaccine may be selected
XX from conserved regions of viral or tumour-associated antigens, which
XX reduces the likelihood of escape mutants, also immunosuppressive epitopes
XX that may be present in whole antigens can be avoided with the use of
XX epitope-base vaccines. An additional advantage is the ability to combine
XX selected epitopes (CTL and HTL) and to modify the composition of the
XX epitopes achieving enhanced immunogenicity, the major benefit of the
XX vaccine is that is safe and efficacious. AAB98391 to AAB98477 represent
XX polypeptide sequences used in the exemplification of the present
XX invention
XX
SQ Sequence 9 AA;
Query Match 72.2%; Score 13; DB 4; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.3e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 YXXXF 5
Db 5 YAAAF 9
RESULT 8
AAB98474
ID AAB98474 standard; protein; 9 AA.
XX
AC AAB98474;
XX

```

DT 22-AUG-2001 (first entry)  
 DE HLA class I standard peptide binding affinity B\*3510.  
 XX Human papillomavirus; human leukocyte antigen; HLA; immune response; HPV;  
 KW epitope; T cell; identification; vaccine; infection; genital wart;  
 KW neoplastic growth; antiviral.  
 XX Homo sapiens.  
 OS WO200141799-A1.  
 PN 14-JUN-2001.  
 PD 11-DEC-2000; 2000WO-US033549.  
 PF 10-DEC-1999; 99US-0172705P.  
 PR 15-AUG-2000; 2000US-00641528.  
 XX (EPIM-) EPIMUNE INC.  
 PA Sette A, Sidney J, Southwood S, Chesnut R, Celis E, Grey HM;  
 XX WPI; 2001-381497/40.  
 DR An isolated human papilloma virus (HPV) epitope, useful in vaccines for  
 XX treating HPV infections.  
 PT Disclosure; Page 98; 756pp; English.  
 PS The present invention describes an isolated prepared human papillomavirus  
 XX (HPV) epitope (I). (I) has antiviral activity, and can be used in vaccine  
 CC production. Peptides and corresponding nucleic acid compositions from the  
 CC present invention are useful for stimulating an immune response to HPV by  
 CC stimulating the production of CTL or HTL responses, specifically in the  
 CC treatment or prophylaxis of HPV infection, in persons who have not  
 CC manifested symptoms e.g. genital warts or neoplastic growth. The peptides  
 CC can also be used in a tetramer staining assay to assess peripheral blood  
 CC mononuclear cells for the presence of antigen-specific CTLs following  
 CC exposure to a pathogen or immunogen, and as reagents to evaluate immune  
 CC recall responses or evaluate the efficacy of a vaccine. The vaccine  
 CC compositions are useful for removing warts or treating HPV infections.  
 CC The epitopes for inclusion in an epitope-base vaccine may be selected  
 CC from conserved regions of viral or tumour-associated antigens, which  
 CC reduces the likelihood of escape mutants, also immunosuppressive epitopes  
 CC that may be present in whole antigens can be avoided with the use of  
 CC epitope-base vaccines. An additional advantage is the ability to combine  
 CC selected epitopes (CTL and HTL) and to modify the composition of the  
 CC epitopes achieving enhanced immunogenicity, the major benefit of the  
 CC vaccine is that is safe and efficacious. AAB98391 to AAB98477 represent  
 CC polypeptide sequences used in the exemplification of the present  
 XX invention  
 XX Sequence 9 AA;  
 SQ Query Match 72.2%; Score 13; DB 4; Length 9;  
 Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 YXXPF 5  
 DB 5 YAAAF 9  
 RESULT 9  
 AAG84528  
 ID AAG84528 standard; peptide; 9 AA.  
 XX AAG84528;  
 AC AAG84528;  
 XX 10-SEP-2001 (first entry)  
 DT Human leukocyte antigen (HLA) class I binding peptide A\*3501.  
 XX

XX Human; human leukocyte antigen; HLA epitope; cytotoxic T lymphocyte; CTL;  
 KW MAGE2; MAGE3; melanoma antigen gene; immune response; vaccine; cancer;  
 KW cytostatic; immunostimulant.  
 XX Homo sapiens.  
 OS Synthetic.  
 OS WO200142267-A1.  
 PN 14-JUN-2001.  
 PD 11-DEC-2000; 2000WO-US033545.  
 PF 10-DEC-1999; 99US-00458298.  
 PR (EPIM-) EPIMUNE INC.  
 PA Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;  
 XX Keogh E;  
 PI WPI; 2001-375002/39.  
 DR An isolated prepared MAGE2/3 epitope (I) for use in pharmaceuticals for  
 XX the treatment and prevention of cancer.  
 PT Disclosure; Page 83; 171pp; English.  
 PS The present invention describes MAGE2/3 epitopes (I). Also described are:  
 CC (1) a clonal cytotoxic T lymphocyte (CTL) that is cultured in vivo and  
 CC binds to a complex of (I); (2) a peptide (II) comprising (I) and a second  
 CC epitope and has less than 50 contiguous amino acids; (3) a vaccine  
 CC composition comprising (II), a unit dose of a peptide with at least 50  
 CC contiguous amino acids with 100% identity to the native peptide sequence  
 CC of MAGE2/3, and a pharmaceutical excipient; (4) an isolated nucleic acid  
 CC encoding (I); and (5) an isolated nucleic acid encoding (II). (I) has  
 CC cytostatic activity, and can be used in vaccines and as an  
 CC immunostimulant. A vaccine of (3) is useful for the treatment and  
 CC prevention of cancer. (I) is useful for monitoring or evaluating an  
 CC immune response by incubating a T-lymphocyte sample from a patient with  
 CC (I) that binds to an human leukocyte antigen (HLA) allele present in the  
 CC patient and detecting the presence of the T-lymphocyte that binds to the  
 CC peptide. The vaccine allows the opportunity to combine epitopes derived  
 CC from multiple tumour-associated molecules reducing the likelihood of  
 CC tumour escape due to antigen loss. AAG84515 to AAG84909 and AAB99725  
 CC represent amino acid sequences used in the exemplification of the present  
 XX invention  
 XX Sequence 9 AA;  
 SQ Query Match 72.2%; Score 13; DB 4; Length 9;  
 Best Local Similarity 40.0%; Pred. No. 1.3e+06;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 YXXPF 5  
 DB 5 YAAAF 9  
 RESULT 10  
 AAU06288  
 ID AAU06288 standard; peptide; 9 AA.  
 XX AAU06288;  
 AC AAU06288;  
 XX 24-OCT-2001 (first entry)  
 DT Human Leukocyte Antigen (HLA) Class I standard supermotif peptide #9.  
 DE Prostate cancer-associated antigen; supermotif; human leukocyte antigen;  
 XX HLA; cytotoxic T lymphocyte; helper T lymphocyte; HLA superfamily; CTL;  
 KW immunogenicity; immunosuppression; HTL.  
 XX





```

PI Baker DM, Celis E, Kubo RT, Grey HM;
XX
DR WPI; 2001-354887/37.
XX
XX Vaccine compositions comprising human immunodeficiency virus-1 (HIV-1)
PT peptide groups, useful for vaccinating against HIV-1.
XX
PS Example 1; Page 416; 448pp; English.
XX
CC The present invention describes a composition (I) comprising a prepared
CC human immunodeficiency virus-1 (HIV-1) group comprising an amino acid
CC sequence selected from 51 defined amino acid sequences (ABL25347 to
CC ABP25397). (I) has virucide activity and can be used in vaccines. (I) may
CC be used for immunising subjects against HIV-1 infections. The use of
CC group-based vaccines has several advantages over traditional vaccines,
CC particularly when compared to the use of whole antigens in vaccine
CC compositions. There is evidence that the immune response to whole
CC antigens is directed largely toward variable regions of the antigen,
CC allowing for immune escape due to mutations. The groups for inclusion in
CC an group-based vaccine may be selected from conserved regions of viral or
CC tumour-associated antigens, which therefore reduces the likelihood of
CC escape mutants. Furthermore, immunosuppressive groups that may be present
CC in whole antigens can be avoided with the use of group-based vaccines. An
CC additional advantage of an group-based vaccine approach is the ability to
CC combine selected groups (CTL and HTL), and further, to modify the
CC composition of the groups, achieving, for example, enhanced
CC immunogenicity. Accordingly, the immune response can be modulated, as
CC appropriate, for the target disease. Similar engineering of the response
CC is not possible with traditional approaches. ABP11501 to ABP25412
CC represent peptide sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 9 AA;
    Query Match      72.2%; Score 13; DB 4; Length 9;
    Best Local Similarity 40.0%; Pred. No. 1.3e+06;
    Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 YXXXF 5
Db 5 YAAAF 9
RESULT 15
ABP11509
ID ABP11509 standard; peptide; 9 AA.
XX
AC ABP11509;
XX
DT 15-JUL-2002 (first entry)
XX
DE HLA class I binding peptide A*3501.
XX
KW HIV; HIV-1; human immunodeficiency virus; env; pol; gag; nef; vpr; vpu;
KW vif; tat; cytotoxic T lymphocyte; CTL; immune response; epitope; antigen;
KW vaccine; HIV infection; immunisation; virucide.
XX
OS Homo sapiens.
OS Synthetic.
XX
FN WO200124810-A1.
XX
PD 12-APR-2001.
XX
XX 05-OCT-2000; 2000WO-US027766.
PF
XX 05-OCT-1999; 99US-00412863.
FR
XX (EPIM-) EPIMMUNE INC.
PA
XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM;
XX

```

---

```

DR WPI; 2001-354887/37.
XX
XX Vaccine compositions comprising human immunodeficiency virus-1 (HIV-1)
PT peptide groups, useful for vaccinating against HIV-1.
XX
PS Example 1; Page 102; 448pp; English.
XX
CC The present invention describes a composition (I) comprising a prepared
CC human immunodeficiency virus-1 (HIV-1) group comprising an amino acid
CC sequence selected from 51 defined amino acid sequences (ABL25347 to
CC ABP25397). (I) has virucide activity and can be used in vaccines. (I) may
CC be used for immunising subjects against HIV-1 infections. The use of
CC group-based vaccines has several advantages over traditional vaccines,
CC particularly when compared to the use of whole antigens in vaccine
CC compositions. There is evidence that the immune response to whole
CC antigens is directed largely toward variable regions of the antigen,
CC allowing for immune escape due to mutations. The groups for inclusion in
CC an group-based vaccine may be selected from conserved regions of viral or
CC tumour-associated antigens, which therefore reduces the likelihood of
CC escape mutants. Furthermore, immunosuppressive groups that may be present
CC in whole antigens can be avoided with the use of group-based vaccines. An
CC additional advantage of an group-based vaccine approach is the ability to
CC combine selected groups (CTL and HTL), and further, to modify the
CC composition of the groups, achieving, for example, enhanced
CC immunogenicity. Accordingly, the immune response can be modulated, as
CC appropriate, for the target disease. Similar engineering of the response
CC is not possible with traditional approaches. ABP11501 to ABP25412
CC represent peptide sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 9 AA;
    Query Match      72.2%; Score 13; DB 4; Length 9;
    Best Local Similarity 40.0%; Pred. No. 1.3e+06;
    Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 YXXXF 5
Db 5 YAAAF 9
Search completed: October 5, 2004, 16:06:45
Job time : 36.9589 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:00:17 ; Search time 9.49315 Seconds  
(without alignments)  
38.068 Million cell updates/sec

Title: US-09-973-473a-28

Perfect score: 18  
Sequence: 1 YXXFX 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	72.2	9	4	US-09-543-608A-50
2	13	72.2	12	2	US-08-811-492-153
3	13	72.2	12	3	US-08-293-728-3
4	13	72.2	12	3	US-09-421-868-3
5	13	72.2	14	1	US-07-841-997A-16
6	13	72.2	14	1	US-08-290-301-16
7	13	72.2	14	4	US-09-013-598-15
8	13	72.2	15	2	US-08-031-538-43
9	13	72.2	15	2	US-08-031-538-56
10	13	72.2	18	1	US-08-401-512-64
11	13	72.2	18	3	US-09-226-012-99
12	13	72.2	28	4	US-09-017-689A-4
13	13	72.2	29	2	US-08-620-151-73
14	13	72.2	29	4	US-09-708-906-7
15	13	72.2	29	4	US-09-708-906-8
16	13	72.2	31	1	US-08-190-802A-239
17	13	72.2	31	3	US-08-477-346-239
18	13	72.2	31	4	US-08-473-089-239
19	13	72.2	31	4	US-08-487-072A-239
20	13	72.2	32	3	US-09-253-396A-214
21	13	72.2	32	4	US-09-708-906-9
22	13	72.2	32	3	US-08-256-104-3
23	13	72.2	37	1	US-08-486-013-16
24	13	72.2	37	2	US-08-482-279-16
25	13	72.2	37	2	US-08-342-268-16
26	13	72.2	37	3	US-09-015-968-16
27	13	72.2	37	4	US-09-397-386-16

28	13	72.2	43	2	US-08-609-046A-2	Sequence 2, Appli
29	13	72.2	43	2	US-08-609-046A-4	Sequence 4, Appli
30	13	72.2	43	3	US-09-158-477-2	Sequence 2, Appli
31	13	72.2	43	3	US-09-158-477-4	Sequence 4, Appli
32	13	72.2	43	4	US-08-858-207A-421	Sequence 421, Appl
33	13	72.2	45	4	US-08-963-851-32	Sequence 32, Appl
34	13	72.2	56	4	US-09-205-258-1193	Sequence 1193, Ap
35	13	72.2	64	3	US-08-303-861-16	Sequence 16, Appl
36	13	72.2	65	4	US-09-107-532A-5828	Sequence 5828, Ap
37	13	72.2	68	4	US-09-107-532A-4746	Sequence 4746, Ap
38	13	72.2	69	4	US-09-252-991A-18048	Sequence 18048, A
39	13	72.2	83	4	US-09-149-476-346	Sequence 346, App
40	13	72.2	96	1	US-08-486-013-21	Sequence 21, Appl
41	13	72.2	96	2	US-08-482-279-21	Sequence 21, Appl
42	13	72.2	96	2	US-08-342-268-21	Sequence 21, Appl
43	13	72.2	96	3	US-09-015-968-21	Sequence 21, Appl
44	13	72.2	96	4	US-09-397-386-21	Sequence 21, Appl
45	13	72.2	97	4	US-09-439-554-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1

US-09-543-608A-50  
; Sequence 50, Application US/09543608A  
; Patent No. 6602510  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John D.  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa A.  
; APPLICANT: Cheamut, Robert  
; APPLICANT: Epimmune Inc  
; TITLE OF INVENTION: HLA Class I A2 Tumor Associated Antigen  
; TITLE OF INVENTION: Peptides and Vaccine Compositions  
; FILE REFERENCE: 018623-015110US  
; CURRENT APPLICATION NUMBER: US/09/543,608A  
; CURRENT FILING DATE: 2002-04-05  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 50  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Standard Peptide 1021.05  
US-09-543-608A-50

Query Match 72.2%; Score 13; DB 4; Length 9;  
Best Local Similarity 40.0%; Pred. No. 2.7e+05;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 YXXFX 5  
Db 5 YAAAF 9

RESULT 2

US-08-811-492-153  
; Sequence 153, Application US/08811492  
; Patent No. 5834247  
; GENERAL INFORMATION:  
; APPLICANT: COMB, DONALD G.  
; APPLICANT: PERLER, FRANCINE B.  
; APPLICANT: JACK, WILLIAM E.  
; APPLICANT: XU, MING-QUN  
; APPLICANT: HODGES, ROBERT A.  
; APPLICANT: NOREN, CHRISTOPHER J.  
; APPLICANT: CHONG, SHAORONG S.C.  
; APPLICANT: ADAM, ERIC

APPLICANT: SOUTHWORTH, MAURICE  
TITLE OF INVENTION: MODIFIED PROTEINS, METHODS OF THEIR  
PRODUCTION AND METHODS FOR PURIFICATION OF TARGET  
PROTEINS

NUMBER OF SEQUENCES: 155  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREGORY D. WILLIAMS; NEW ENGLAND BIOLABS, INC.  
STREET: 32 TOZER ROAD  
CITY: BEVERLY  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 01915

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC\ DOS\MS\ DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,492  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/580,555  
FILING DATE: 29-DEC-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/496,247  
FILING DATE: 28-JUN-1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/146,885  
FILING DATE: 03-NOV-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/004,139  
FILING DATE: 09-DEC-1992  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Williams, Gregory D  
REGISTRATION NUMBER: 30901  
REFERENCE/DOCKET NUMBER: NEB-036C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 508-927-5054  
TELEFAX: 509-927-1705  
TELEX:

INFORMATION FOR SEQ ID NO: 153:

SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acid  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-811-492-153

Query Match 72.2%; Score 13; DB 2; Length 12;  
Best Local Similarity 40.0%; Pred. No. 4.7e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
DB 2 YASTF 6

## RESULT 3

US-08-293-728-3  
Sequence 3, Application US/08293728D  
Patent No. 6008341

GENERAL INFORMATION:  
APPLICANT: Foster, Timothy J.  
TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene  
FILE REFERENCE: 05344.105011  
CURRENT APPLICATION NUMBER: US/08/293,728D

CURRENT FILING DATE: 1994-08-22  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-08-293-728-3

Query Match 72.2%; Score 13; DB 3; Length 12;  
Best Local Similarity 40.0%; Pred. No. 4.7e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
DB 5 YTSAP 9

## RESULT 4

US-09-421-868-3  
Sequence 3, Application US/09421868  
Patent No. 6177084

GENERAL INFORMATION:  
APPLICANT: Foster, Timothy J.  
TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene  
FILE REFERENCE: 05344.105011  
CURRENT APPLICATION NUMBER: US/09/421,868  
CURRENT FILING DATE: 1999-10-19  
PRIOR APPLICATION NUMBER: 08/293,728  
PRIOR FILING DATE: 1994-08-22  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-421-868-3

Query Match 72.2%; Score 13; DB 3; Length 12;  
Best Local Similarity 40.0%; Pred. No. 4.7e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
DB 5 YTSAP 9

## RESULT 5

US-07-841-997A-16  
Sequence 16, Application US/07841997A  
Patent No. 5422254

GENERAL INFORMATION:  
APPLICANT: Londerborough, John  
TITLE OF INVENTION: A method to increase the trehalose content  
of organisms by transforming them with the  
TITLE OF INVENTION: structural genes for the short and long chains  
TITLE OF INVENTION: yeast trehalose synthase.  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Alko Ltd.  
STREET: PO Box 350  
CITY: Helsinki  
STATE: -  
COUNTRY: Finland  
ZIP: SF-00101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
COMPUTER: IBM PC/XT/AT  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WP5.1 file exported as DOS text file  
CURRENT APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:  
NAME: Kubovcik, Ronald J.  
FILING DATE: 19920228  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/836,021  
FILING DATE: February 14, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary E. Gormley  
REGISTRATION NUMBER: 34409  
REFERENCE/DOCKET NUMBER: 920085A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)659-2930  
TELEFAX: (202)887-0357  
TELEX: 440142  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-07-841-997A-16

Query Match 72.2%; Score 13; DB 1; Length 14;  
Best Local Similarity 40.0%; Pred. No. 5.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXF 5  
Db 1 YTSAP 5

RESULT 6  
US-08-290-301-16  
Sequence 16, Application US/08290301  
Patent No. 5792921  
GENERAL INFORMATION:  
APPLICANT: Lonsborough, John  
APPLICANT: Tunnela, Outi  
APPLICANT: Palva, Tupio  
APPLICANT: Holmstrom, Kjell-Ove  
APPLICANT: Wellin, Bjorn  
APPLICANT: Mandel, Abul  
TITLE OF INVENTION: Increasing the trehalose content  
of organisms by transforming them with combinations of  
the structural genes for trehalose synthase.  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Alko Ltd.  
STREET: PO Box 350  
CITY: Helsinki  
STATE: Finland  
COUNTRY: Finland  
ZIP: SF-00101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
COMPUTER: IBM PC/XT/AT  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WP5.1 file exported as DOS text file  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,301  
FILING DATE: 15 August 1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FI 943133  
FILING DATE: 29 June 1994  
APPLICATION NUMBER: PCT/FI93/00049  
FILING DATE: 15 February 1993  
APPLICATION NUMBER: 07/841,997  
FILING DATE: 28 February 1992  
APPLICATION NUMBER: 07/836,021  
FILING DATE: 14 February 1992

ATTORNEY/AGENT INFORMATION:  
NAME: Kubovcik, Ronald J.  
FILING DATE: 19920228  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/836,021  
FILING DATE: February 14, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary E. Gormley  
REGISTRATION NUMBER: 34409  
REFERENCE/DOCKET NUMBER: 920085A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)659-2930  
TELEFAX: (202)887-0357  
TELEX: 440142  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-290-301-16

Query Match 72.2%; Score 13; DB 1; Length 14;  
Best Local Similarity 40.0%; Pred. No. 5.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXF 5  
Db 1 YTSAP 5

RESULT 7  
US-09-013-598-16  
Sequence 16, Application US/09013598  
Patent No. 6323001  
GENERAL INFORMATION:  
APPLICANT: Lonsborough, John  
APPLICANT: Tunnela, Outi  
APPLICANT: Palva, Tupio  
APPLICANT: Holmstrom, Kjell-Ove  
APPLICANT: Wellin, Bjorn  
APPLICANT: Mandel, Abul  
TITLE OF INVENTION: Increasing the trehalose content  
of organisms by transforming them with combinations of  
the structural genes for trehalose synthase.  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Alko Ltd.  
STREET: PO Box 350  
CITY: Helsinki  
STATE: Finland  
COUNTRY: Finland  
ZIP: SF-00101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
COMPUTER: IBM PC/XT/AT  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WP5.1 file exported as DOS text file  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/013,598  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/290,301  
FILING DATE:  
APPLICATION NUMBER: PCT/FI93/00049  
FILING DATE: 15 February 1993  
APPLICATION NUMBER: 07/841,997  
FILING DATE: 28 February 1992  
APPLICATION NUMBER: 07/836,021  
FILING DATE: 14 February 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kubovcik, Ronald J.  
FILING DATE: 19920228  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/836,021  
FILING DATE: 14 February 1992

REFERENCE/DOCKET NUMBER: LAIN-001  
TELEPHONE: (202) 467-6300  
TELEFAX: (202) 466-2006  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-09-013-598-16

Query Match 72.2%; Score 13; DB 4; Length 14;  
Best Local Similarity 40.0%; Pred. No. 5.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
DB 1 YTSAP 5

RESULT 8  
US-08-031-538-43  
Sequence 43, Application US/08031538  
Patent No. 5968817  
GENERAL INFORMATION:  
APPLICANT: Sutcliffe, J Gregor  
APPLICANT: Erlander, Mark G  
APPLICANT: Lovenberg, Timothy W  
TITLE OF INVENTION: HUMAN SEROTONIN RECEPTORS, DNA ENCODING  
TITLE OF INVENTION: THE RECEPTORS, AND USES THEREOF  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10666 No. 5968817th Torrey Pines Road, TPC 8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 19930315  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSP5099P  
TELEPHONE: 619-554-6312  
TELEFAX: 619-554-2937  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-031-538-43

Query Match 72.2%; Score 13; DB 2; Length 15;  
Best Local Similarity 40.0%; Pred. No. 5.5e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
DB 3 YSSAF 7

RESULT 9  
US-08-031-538-56  
Sequence 56, Application US/08031538  
Patent No. 5968817  
GENERAL INFORMATION:  
APPLICANT: Sutcliffe, J Gregor  
APPLICANT: Erlander, Mark G  
APPLICANT: Lovenberg, Timothy W  
TITLE OF INVENTION: HUMAN SEROTONIN RECEPTORS, DNA ENCODING  
TITLE OF INVENTION: THE RECEPTORS, AND USES THEREOF  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10666 No. 5968817th Torrey Pines Road, TPC 8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 19930315  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSP5099P  
TELEPHONE: 619-554-2937  
TELEFAX: 619-554-6312  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-031-538-56

Query Match 72.2%; Score 13; DB 2; Length 15;  
Best Local Similarity 40.0%; Pred. No. 5.5e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
DB 3 YSSAF 7

RESULT 10  
US-08-401-512-64  
Sequence 64, Application US/08401512  
Patent No. 5599673  
GENERAL INFORMATION:  
APPLICANT: Keating, Mark T.  
APPLICANT: Curran, Mark E.  
APPLICANT: Wang, Qing  
TITLE OF INVENTION: Long QT Syndrome Genes  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3917  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/401,512  
; FILING DATE: 09-MAR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saxe, Stephen A.  
; REGISTRATION NUMBER: 38,609  
; REFERENCE/DOCKET NUMBER: 19780-113879  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-962-4848  
; TELEFAX: 202-962-8300  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-08-401-512-64

Query Match 72.2%; Score 13; DB 1; Length 18;  
Best Local Similarity 40.0%; Pred. No. 6.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
DB 10 YSAAF 14

RESULT 11  
US-09-226-012-99  
; Sequence 99, Application US/09226012  
; Patent No. 6207383  
; GENERAL INFORMATION:  
; APPLICANT: Keating, Mark T.  
; APPLICANT: Splawski, Igor  
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT  
; FILE REFERENCE: 2323-136  
; CURRENT APPLICATION NUMBER: US/09/226,012  
; CURRENT FILING DATE: 1999-01-06  
; EARLIER APPLICATION NUMBER: 09/122,847  
; EARLIER FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 99  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-226-012-99

Query Match 72.2%; Score 13; DB 3; Length 18;  
Best Local Similarity 40.0%; Pred. No. 6.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
DB 10 YSAAF 14

RESULT 12  
US-09-017-689A-4  
; Sequence 4, Application US/09017689A  
; Patent No. 6413940  
; GENERAL INFORMATION:  
; APPLICANT: AVERBACK, Paul  
; TITLE OF INVENTION: PHARMACEUTICALLY ACTIVE AGENTS THAT  
; IMPED THE FORMATION OF AMYLOID BY IMPEDING THE GENESIS OF

; DMS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/017,689A  
; FILING DATE: 03-Feb-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/038,694  
; FILING DATE: 07-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 018792/0125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-017-689A-4  
  
Query Match 72.2%; Score 13; DB 4; Length 28;  
Best Local Similarity 40.0%; Pred. No. 8.4e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 YXXXF 5  
DB 23 YATTF 27  
  
RESULT 13  
US-08-620-151-73  
; Sequence 73, Application US/08620151  
; Patent No. 5928955  
; GENERAL INFORMATION:  
; APPLICANT: Imperiali, Barbara  
; APPLICANT: Walkup, Grant K.  
; TITLE OF INVENTION: PEPTIDYL FLUORESCENT CHEMOSENSOR FOR  
; NUMBER OF SEQUENCES: 136  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BRINKS, HOFER, GILSON & LIONE  
; STREET: NBC Tower - Suite 3600, 455 N. Cityfront  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60611-5599  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/620,151  
; FILING DATE: 22-MAR-1996

```
; CLASSIFICATION: 422
; ATTORNEY/AGENT INFORMATION:
; NAME: Shannon, Karen L.
; REGISTRATION NUMBER: 36,675
; REFERENCE/DOCKET NUMBER: 8597/6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-321-4200
; TELEFAX: 312-321-4299
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-620-151-73

Query Match 72.2%; Score 13; DB 2; Length 29;
Best Local Similarity 40.0%; Pred. No. 8.6e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
Db 13 YTTAF 17

RESULT 14
US-09-708-906-7
; Sequence 7, Application US/09708906
; Patent No. 6528620
; GENERAL INFORMATION:
; APPLICANT: Ayer, Donald E.
; APPLICANT: Billin, Andrew N.
; TITLE OF INVENTION: SID-POLYAMIDE FUSIONS: A POTENT METHOD OF REGULATING
; TITLE OF INVENTION: GENE EXPRESSION
; FILE REFERENCE: 1321.2.37
; CURRENT APPLICATION NUMBER: US/09/708,906
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/163,960
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-708-906-7

Query Match 72.2%; Score 13; DB 4; Length 29;
Best Local Similarity 40.0%; Pred. No. 8.6e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
Db 24 YASSF 28

RESULT 15
US-09-708-906-8
; Sequence 8, Application US/09708906
; Patent No. 6528620
; GENERAL INFORMATION:
; APPLICANT: Ayer, Donald E.
; APPLICANT: Billin, Andrew N.
; TITLE OF INVENTION: SID-POLYAMIDE FUSIONS: A POTENT METHOD OF REGULATING
; TITLE OF INVENTION: GENE EXPRESSION
; FILE REFERENCE: 1321.2.37
; CURRENT APPLICATION NUMBER: US/09/708,906
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/163,960
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 8
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-708-906-8

Query Match 72.2%; Score 13; DB 4; Length 29;
Best Local Similarity 40.0%; Pred. No. 8.6e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
Db 24 YASSF 28

Search completed: October 5, 2004, 16:15:42
Job time : 10.4932 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 16:12:48 ; Search time 33.7534 Seconds  
(without alignments)  
66.737 Million cell updates/sec

Title: US-09-973-473A-28  
Perfect score: 18  
Sequence: 1 YXXFX 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	72.2	6	US-10-174-209-70	Sequence 70, Appl
2	13	72.2	7	US-10-327-598-506	Sequence 506, App
3	13	72.2	9	US-08-452-843A-1	Sequence 1, Appl
4	13	72.2	9	US-08-344-824-17	Sequence 17, Appl
5	13	72.2	9	US-09-977-797A-50	Sequence 50, Appl
6	13	72.2	9	US-10-149-135-2190	Sequence 2190, Ap
7	13	72.2	9	US-10-149-135-2191	Sequence 2191, Ap
8	13	72.2	9	US-10-149-135-2192	Sequence 2192, Ap
9	13	72.2	9	US-10-149-135-2193	Sequence 2193, Ap
10	13	72.2	9	US-10-149-135-2305	Sequence 2305, Ap
11	13	72.2	9	US-10-149-135-2306	Sequence 2306, Ap
12	13	72.2	9	US-10-149-135-2307	Sequence 2307, Ap
13	13	72.2	9	US-10-149-135-2309	Sequence 2309, Ap
14	13	72.2	9	US-10-149-135-2310	Sequence 2310, Ap
15	13	72.2	9	US-10-149-135-2311	Sequence 2311, Ap

16	13	72.2	9	15	US-10-149-138-4203	Sequence 4203, Ap
17	13	72.2	9	15	US-10-149-138-4204	Sequence 4204, Ap
18	13	72.2	9	15	US-10-149-138-4205	Sequence 4205, Ap
19	13	72.2	9	15	US-10-149-138-4206	Sequence 4206, Ap
20	13	72.2	9	15	US-10-149-138-4365	Sequence 4365, Ap
21	13	72.2	9	15	US-10-149-138-4366	Sequence 4366, Ap
22	13	72.2	9	15	US-10-149-138-4367	Sequence 4367, Ap
23	13	72.2	9	15	US-10-149-138-4369	Sequence 4369, Ap
24	13	72.2	9	15	US-10-149-138-4370	Sequence 4370, Ap
25	13	72.2	9	15	US-10-149-138-4371	Sequence 4371, Ap
26	13	72.2	9	16	US-10-149-138-4203	Sequence 4203, Ap
27	13	72.2	9	16	US-10-149-138-4204	Sequence 4204, Ap
28	13	72.2	9	16	US-10-149-138-4205	Sequence 4205, Ap
29	13	72.2	9	16	US-10-149-138-4206	Sequence 4206, Ap
30	13	72.2	9	16	US-10-149-138-4365	Sequence 4365, Ap
31	13	72.2	9	16	US-10-149-138-4366	Sequence 4366, Ap
32	13	72.2	9	16	US-10-149-138-4367	Sequence 4367, Ap
33	13	72.2	9	16	US-10-149-138-4369	Sequence 4369, Ap
34	13	72.2	9	16	US-10-149-138-4370	Sequence 4370, Ap
35	13	72.2	9	16	US-10-149-138-4371	Sequence 4371, Ap
36	13	72.2	13	9	US-09-982-172-193	Sequence 193, App
37	13	72.2	13	9	US-09-982-172-246	Sequence 246, App
38	13	72.2	13	12	US-10-267-565-16	Sequence 16, Appl
39	13	72.2	13	14	US-10-349-507-16	Sequence 16, Appl
40	13	72.2	14	14	US-10-174-209-77	Sequence 77, Appl
41	13	72.2	15	16	US-10-203-915A-173	Sequence 173, App
42	13	72.2	15	16	US-10-203-915A-174	Sequence 174, App
43	13	72.2	15	16	US-10-203-915A-175	Sequence 175, App
44	13	72.2	16	11	US-09-791-551-15	Sequence 15, Appl
45	13	72.2	18	9	US-09-735-995-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1

US-10-174-209-70  
; Sequence 70, Application US/10174209  
; Publication No. US20030177526A1  
; GENERAL INFORMATION:  
; APPLICANT: Song, Xiaoling  
; APPLICANT: Bariola, Pauline A.  
; APPLICANT: Linderoth, No. US20030177526A1a A.  
; APPLICANT: Fan, Hao  
; APPLICANT: Wei, Zhong-Min  
; TITLE OF INVENTION: RECEPTORS FOR HYPERSENSITIVE RESPONSE ELICITORS AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: 21829/211  
; CURRENT APPLICATION NUMBER: US/10/174,209  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: 60/335,776  
; PRIOR FILING DATE: 2001-10-31  
; PRIOR APPLICATION NUMBER: 09/810,997  
; PRIOR FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 86  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 70  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: peptide  
US-10-174-209-70

Query Match 72.2%; Score 13; DB 14; Length 6;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXFX 5

Db 1 YSSAF 5

```
RESULT 2
US-10-327-598-506
; Sequence 506, Application US/10327598
; Publication No. US20040181039A1
; GENERAL INFORMATION:
; APPLICANT: Krah, Eugene
; APPLICANT: Guo, Hongliang
; APPLICANT: Aiyappa, Ashok
; APPLICANT: Lawton, Robert
; TITLE OF INVENTION: Canine Immunoglobulin Variable Domains, Caninized Antibodies, and
; TITLE OF INVENTION: for Making and Using Them
; FILE REFERENCE: 01-799-A
; CURRENT APPLICATION NUMBER: US/10/327,598
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/344,874
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 1139
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 506
; LENGTH: 7
; TYPE: PRT
; ORGANISM: canis familiaris;
US-10-327-598-506

Query Match      72.2%; Score 13; DB 16; Length 7;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
Db 2 YSSSF 6

RESULT 3
US-08-452-843A-1
; Sequence 1, Application US/08452843A
; Publication No. US20020098197A1
; GENERAL INFORMATION:
; APPLICANT: Sette, Alessandro
; APPLICANT: Sidney, John
; TITLE OF INVENTION: HLA Binding Peptides and Their Uses
; FILE REFERENCE: 399632001321
; CURRENT APPLICATION NUMBER: US/08/452,843A
; CURRENT FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: US 08/344,824
; PRIOR FILING DATE: 1994-11-23
; PRIOR APPLICATION NUMBER: US 08/278,634
; PRIOR FILING DATE: 1994-07-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: B35 consensus peptide
US-08-452-843A-1

Query Match      72.2%; Score 13; DB 8; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
Db 5 YAAAF 9

RESULT 4
US-08-344-824-17
; Sequence 17, Application US/08344824
; Publication No. US20030152580A1
; GENERAL INFORMATION:
; APPLICANT: SETTE, Alessandro
; APPLICANT: SETTE, Alessandro
```

```
; APPLICANT: SIDNEY, John
; TITLE OF INVENTION: HLA BINDING PEPTIDES AND THEIR USES
; NUMBER OF SEQUENCES: 399
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/344,824
; FILING DATE: 23-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,634
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 14137-80-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-344-824-17

Query Match      72.2%; Score 13; DB 8; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
Db 5 YAAAF 9

RESULT 5
US-09-977-797A-50
; Sequence 50, Application US/09977797A
; Publication No. US20030044772A1
; GENERAL INFORMATION:
; APPLICANT: Watkins, Jeffrey D.
; APPLICANT: Huse, William D.
; APPLICANT: Wu, Herren
; TITLE OF INVENTION: Methods for Identifying Ligand Specific Binding Molecules
; FILE REFERENCE: AME-06805
; CURRENT APPLICATION NUMBER: US/09/977,797A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 09/129,026
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 08/905,825
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 136
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 50
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-797A-50

Query Match      72.2%; Score 13; DB 10; Length 9;
```



```
Best Local Similarity 40.0%; Pred. No. 1.2e+06; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 3;

Qy 1 YXXXF 5
Db 3 YSSTF 7

RESULT 6
US-10-149-135-2190
; Sequence 2190, Application US/10149135
; Publication No. US20040053822A1
; GENERAL INFORMATION:
; APPLICANT: Fikes, John
; APPLICANT: Sette, Alessandro
; APPLICANT: Sidney, John
; APPLICANT: Southwood, Scott
; APPLICANT: Chesnut, Robert
; APPLICANT: Celis, Esteban
; APPLICANT: Keogh, Elissa
; TITLE OF INVENTION: Inducing Cellular Immune Responses to
; FILE REFERENCE: 2060.0130001
; CURRENT APPLICATION NUMBER: US/10/149,135
; CURRENT FILING DATE: 2000-12-11
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 09/458,298
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-06-04
; PRIOR APPLICATION NUMBER: US 08/027,146
; NUMBER OF SEQ ID NOS: 2479
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2190
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificial Peptide
US-10-149-135-2190

Query Match 72.2%; Score 13; DB 12; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXF 5
Db 5 YAAAF 9

RESULT 7
US-10-149-135-2191
; Sequence 2191, Application US/10149135
; Publication No. US20040053822A1
; GENERAL INFORMATION:
; APPLICANT: Fikes, John
; APPLICANT: Sette, Alessandro
; APPLICANT: Sidney, John
; APPLICANT: Southwood, Scott
; APPLICANT: Chesnut, Robert
; APPLICANT: Celis, Esteban
; APPLICANT: Keogh, Elissa
; TITLE OF INVENTION: Inducing Cellular Immune Responses to
; FILE REFERENCE: 2060.0130001
; CURRENT APPLICATION NUMBER: US/10/149,135
; CURRENT FILING DATE: 2000-12-11
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 09/458,298
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-06-04
; PRIOR APPLICATION NUMBER: US 08/027,146
; NUMBER OF SEQ ID NOS: 2479
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2191
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificial Peptide
US-10-149-135-2190

Query Match 72.2%; Score 13; DB 12; Length 9;
Best Local Similarity 40.0%; Pred. No. 1.2e+06;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXF 5
Db 5 YAAAF 9

RESULT 8
US-10-149-135-2192
; Sequence 2192, Application US/10149135
; Publication No. US20040053822A1
; GENERAL INFORMATION:
; APPLICANT: Fikes, John
; APPLICANT: Sette, Alessandro
; APPLICANT: Sidney, John
; APPLICANT: Southwood, Scott
; APPLICANT: Chesnut, Robert
; APPLICANT: Celis, Esteban
; APPLICANT: Keogh, Elissa
; TITLE OF INVENTION: Inducing Cellular Immune Responses to
; FILE REFERENCE: 2060.0130001
; CURRENT APPLICATION NUMBER: US/10/149,135
; CURRENT FILING DATE: 2000-12-11
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 09/458,298
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-06-04
; PRIOR APPLICATION NUMBER: US 08/027,146
; NUMBER OF SEQ ID NOS: 2479
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2192
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificial Peptide
US-10-149-135-2192
```

; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2192

Query Match 72.2%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
Db 5 YAAAF 9

RESULT 9

US-10-149-135-2193  
; Sequence 2193, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE OF INVENTION: MAGE2/3 Using Peptide and Nucleic Acid Compositions  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; PRIOR FILING DATE: 1993-03-05  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2193  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2193

Query Match 72.2%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
Db 5 YAAAF 9

RESULT 10

US-10-149-135-2305  
; Sequence 2305, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban

; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE OF INVENTION: MAGE2/3 Using Peptide and Nucleic Acid Compositions  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; PRIOR FILING DATE: 1993-03-05  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2305  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2305

Query Match 72.2%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
Db 5 YAAAF 9

RESULT 11

US-10-149-135-2306  
; Sequence 2306, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; FILE OF INVENTION: MAGE2/3 Using Peptide and Nucleic Acid Compositions  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; PRIOR FILING DATE: 1993-03-05  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2306

LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2306

Query Match 72.2%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
Db 5 YAAAF 9

RESULT 12  
US-10-149-135-2307  
Sequence 2307, Application US/10149135  
Publication No. US20040053822A1  
GENERAL INFORMATION:  
APPLICANT: Fikes, John  
APPLICANT: Sette, Alessandro  
APPLICANT: Sidney, John  
APPLICANT: Southwood, Scott  
APPLICANT: Chesnut, Robert  
APPLICANT: Celis, Esteban  
APPLICANT: Keogh, Elissa  
TITLE OF INVENTION: Inducing Cellular Immune Responses to  
TITLE OF INVENTION: MAGE2/3 Using Peptide and Nucleic Acid Compositions  
FILE REFERENCE: 2060.0130001  
CURRENT APPLICATION NUMBER: US/10/149,135  
CURRENT FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: PCT/US00/33545  
PRIOR FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: US 09/458,298  
PRIOR FILING DATE: 1999-12-10  
PRIOR APPLICATION NUMBER: US 09/189,702  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: US 08/205,713  
PRIOR FILING DATE: 1994-03-04  
PRIOR APPLICATION NUMBER: US 08/159,184  
PRIOR FILING DATE: 1993-11-29  
PRIOR APPLICATION NUMBER: US 08/073,205  
PRIOR FILING DATE: 1993-06-04  
PRIOR APPLICATION NUMBER: US 08/027,146  
PRIOR FILING DATE: 1993-03-05  
NUMBER OF SEQ ID NOS: 2479  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 2307  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2307

Query Match 72.2%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
Db 5 YAAAF 9

RESULT 13  
US-10-149-135-2309  
Sequence 2309, Application US/10149135  
Publication No. US20040053822A1  
GENERAL INFORMATION:  
APPLICANT: Fikes, John  
APPLICANT: Sette, Alessandro

APPLICANT: Sidney, John  
APPLICANT: Southwood, Scott  
APPLICANT: Chesnut, Robert  
APPLICANT: Celis, Esteban  
APPLICANT: Keogh, Elissa  
TITLE OF INVENTION: Inducing Cellular Immune Responses to  
TITLE OF INVENTION: MAGE2/3 Using Peptide and Nucleic Acid Compositions  
FILE REFERENCE: 2060.0130001  
CURRENT APPLICATION NUMBER: US/10/149,135  
CURRENT FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: PCT/US00/33545  
PRIOR FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: US 09/458,298  
PRIOR FILING DATE: 1999-12-10  
PRIOR APPLICATION NUMBER: US 09/189,702  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: US 08/205,713  
PRIOR FILING DATE: 1994-03-04  
PRIOR APPLICATION NUMBER: US 08/159,184  
PRIOR FILING DATE: 1993-11-29  
PRIOR APPLICATION NUMBER: US 08/073,205  
PRIOR FILING DATE: 1993-06-04  
PRIOR APPLICATION NUMBER: US 08/027,146  
PRIOR FILING DATE: 1993-03-05  
NUMBER OF SEQ ID NOS: 2479  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 2309  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2309

Query Match 72.2%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
| |  
Db 5 YAAAF 9

RESULT 14  
US-10-149-135-2310  
Sequence 2310, Application US/10149135  
Publication No. US20040053822A1  
GENERAL INFORMATION:  
APPLICANT: Fikes, John  
APPLICANT: Sette, Alessandro  
APPLICANT: Sidney, John  
APPLICANT: Southwood, Scott  
APPLICANT: Chesnut, Robert  
APPLICANT: Celis, Esteban  
APPLICANT: Keogh, Elissa  
TITLE OF INVENTION: Inducing Cellular Immune Responses to  
TITLE OF INVENTION: MAGE2/3 Using Peptide and Nucleic Acid Compositions  
FILE REFERENCE: 2060.0130001  
CURRENT APPLICATION NUMBER: US/10/149,135  
CURRENT FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: PCT/US00/33545  
PRIOR FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: US 09/458,298  
PRIOR FILING DATE: 1999-12-10  
PRIOR APPLICATION NUMBER: US 09/189,702  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: US 08/205,713  
PRIOR FILING DATE: 1994-03-04  
PRIOR APPLICATION NUMBER: US 08/159,184  
PRIOR FILING DATE: 1993-11-29  
PRIOR APPLICATION NUMBER: US 08/073,205  
PRIOR FILING DATE: 1993-06-04  
PRIOR APPLICATION NUMBER: US 08/027,146

; PRIOR FILING DATE: 1993-03-05  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2310  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2310

Query Match 72.2%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. NO. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXP 5  
Db 5 YAAAF 9

RESULT 15  
US-10-149-135-2311  
; Sequence 2311, Application US/10149135  
; Publication No. US20040053822A1  
; GENERAL INFORMATION:  
; APPLICANT: Fikes, John  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Sidney, John  
; APPLICANT: Southwood, Scott  
; APPLICANT: Chesnut, Robert  
; APPLICANT: Celis, Esteban  
; APPLICANT: Keogh, Elissa  
; TITLE OF INVENTION: Inducing Cellular Immune Responses to  
; TITLE OF INVENTION: MAGE2/3 Using Peptide and Nucleic Acid Compositions  
; FILE REFERENCE: 2060.0130001  
; CURRENT APPLICATION NUMBER: US/10/149,135  
; CURRENT FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: PCT/US00/33545  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 09/458,298  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: US 09/189,702  
; PRIOR FILING DATE: 1998-11-10  
; PRIOR APPLICATION NUMBER: US 08/205,713  
; PRIOR FILING DATE: 1994-03-04  
; PRIOR APPLICATION NUMBER: US 08/159,184  
; PRIOR FILING DATE: 1993-11-29  
; PRIOR APPLICATION NUMBER: US 08/073,205  
; PRIOR FILING DATE: 1993-06-04  
; PRIOR APPLICATION NUMBER: US 08/027,146  
; PRIOR FILING DATE: 1993-03-05  
; NUMBER OF SEQ ID NOS: 2479  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2311  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificial Peptide  
US-10-149-135-2311

Query Match 72.2%; Score 13; DB 12; Length 9;  
Best Local Similarity 40.0%; Pred. NO. 1.2e+06;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXP 5  
Db 5 YAAAF 9

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:58:01 ; Search time 8.53425 Seconds  
(without alignments)  
78.899 Million cell updates/sec

Title: US-09-973-473a-28

Perfect score: 18

Sequence: 1 YXXFX 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	72.2	38	2 S22210	photosystem I prot
2	13	72.2	40	2 H95063	hypothetical prote
3	13	72.2	46	2 S17919	photosystem II pro
4	13	72.2	57	2 G97910	hypothetical prote
5	13	72.2	58	2 C82818	hypothetical prote
6	13	72.2	68	2 G82600	hypothetical prote
7	13	72.2	70	2 I77547	hypothetical prote
8	13	72.2	72	2 G89851	hypothetical prote
9	13	72.2	75	2 H97803	proline/betaine tr
10	13	72.2	77	2 E49786	bacteriocin probab
11	13	72.2	79	2 AD1761	B. subtilis CsbA p
12	13	72.2	84	2 JT0457	hypothetical prote
13	13	72.2	87	2 T03993	hypothetical prote
14	13	72.2	90	2 C81917	hypothetical prote
15	13	72.2	90	2 A81186	hypothetical prote
16	13	72.2	93	2 S27028	hypothetical prote
17	13	72.2	101	2 A13460	bacterial protein
18	13	72.2	102	2 PH1254	Ig heavy chain v r
19	13	72.2	108	2 S51958	hypothetical prote
20	13	72.2	108	2 H90122	dna directed RNA p
21	13	72.2	108	2 G72635	hypothetical prote
22	13	72.2	112	2 S58139	gene 9 protein - p
23	13	72.2	112	2 S47298	sulfolysin - Strepto
24	13	72.2	115	2 G71058	hypothetical prote
25	13	72.2	115	2 T38415	very hypothetical
26	13	72.2	116	2 S22553	Ig heavy chain v r
27	13	72.2	117	2 F87673	conserved hypothet
28	13	72.2	119	2 H71035	hypothetical prote
29	13	72.2	119	2 H84991	hypothetical prote

30 13 72.2 119 2 B90074 hypothetical prote  
31 13 72.2 119 2 AG3219 hypothetical prote  
32 13 72.2 120 2 B72519 hypothetical prote  
33 13 72.2 121 2 B70767 hypothetical prote  
34 13 72.2 122 2 D89803 conserved hypothet  
35 13 72.2 122 2 AD0327 probable exported  
36 13 72.2 128 2 T41487 very hypothetical  
37 13 72.2 129 2 H90324 conserved hypothet  
38 13 72.2 130 2 E90425 conserved hypothet  
39 13 72.2 131 2 H87452 conserved hypothet  
40 13 72.2 132 2 A90731 hypothetical prote  
41 13 72.2 134 2 S49531 anti-Sm antibody V  
42 13 72.2 134 2 T20516 hypothetical prote  
43 13 72.2 134 2 A88691 protein F4H10.9  
44 13 72.2 134 2 H82969 hypothetical prote  
45 13 72.2 135 2 T01620 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S22210  
photosystem I protein psal - Synechococcus sp.

N:Alternate names: photosystem I chain VIII

C:Species: Synechococcus sp.

C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Mar-2001

C:Accession: S22210

R:Muehlenhoff, U.; Haehnel, W.; Witt, H.T.; Herrmann, R.G.

submitted to the EMBL Data Library, January 1992

A:Description: Genes encoding ten subunits of photosystem I from the thermophilic cyanob

A:Reference number: S18970

A:Accession: S22210

A:Molecule type: DNA

A:Residues: 1-38 <MUE>

A:Cross-references: EMBL:X63763; NID:G47585; PIDN:CAA45297.1; PID:G47588

C:Genetics:

A:Gene: psal

C:Superfamily: photosystem I protein psal

C:Keywords: membrane-associated complex; photosynthesis; photosystem I; thylakoid

Query Match 72.2%; Score 13; DB 2; Length 38;

Best Local Similarity 40.0%; Pred. No. 4.7e+02;

Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXFX 5

Db 5 YAAAF 9

##### RESULT 2

H95063

hypothetical protein SP0548 [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae,

C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: H95063

R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide

on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Lofsky, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: H95063

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-40 <KUR>

A:Cross-references: GB:AB005672; PIDN:AAK74705.1; PID:G14972023; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0548

Query Match 72.2%; Score 13; DB 2; Length 40;

Best Local Similarity 40.0%; Pred. No. 4.9e+02;		Mismatches 3; Indels 0; Gaps 0;	
Matches 2; Conservative 0;			
QY	1 YXXXF 5		
Db	29 YSTSF 33		
RESULT 3			
S17919			
photosystem II protein psbK - garden pea chloroplast (fragment)			
C;Species: chloroplast Pisum sativum (garden pea)			
C;Date: 13-Jan-1995 #sequence_revision 10-Oct-1997 #text_change 26-Aug-1999			
C;Accession: S17919; S39471			
R;Nagano, Y.; Matsuno, R.; Sasaki, Y.			
Curr. Genet. 20, 431-436, 1991			
A;Title: Sequence and transcriptional analysis of the gene cluster trnQ-zfpA-psaI-ORF231			
A;Reference number: S17919; MUID:92224289; PMID:1807835			
A;Accession: S17919			
A;Molecule type: DNA			
A;Residues: 1-29 <NAG>			
A;Cross-references: EMBL:X56315; NID:g12187; PIDN:CAA39753.1; PID:g12188			
R;Zakharov, S.D.; Ewy, R.G.; Dilley, R.A.			
FEBS Lett. 336, 95-99, 1993			
A;Title: Subunit III of the chloroplast ATP-synthase can form a Ca(2+)-binding site on t			
A;Reference number: S39470; MUID:94085601; PMID:8262226			
A;Accession: S39471			
A;Molecule type: protein			
A;Residues: 25-46 <ZAK>			
C;Genetics:			
A;Gene: psbK			
A;Genome: chloroplast			
C;Superfamily: photosystem II protein psbK			
C;Keywords: chloroplast; membrane-associated complex; photosynthesis; photosystem II; th			
Query Match 72.2%; Score 13; DB 2; Length 46;			
Best Local Similarity 40.0%; Pred. No. 5.4e+02;			
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1 YXXXF 5		
Db	17 YSSSF 21		
RESULT 4			
G97910			
hypothetical protein spr0311 [imported] - Streptococcus pneumoniae (strain R6)			
C;Species: Streptococcus pneumoniae			
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001			
C;Accession: G97910			
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; H			
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M			
Y, P.; Sun, P.M.; Winkler, M.E.			
J. Bacteriol. 183, 5709-5717, 2001			
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.			
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.			
A;Reference number: A57872; MUID:21429245; PMID:11544234			
A;Accession: G97910			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-57 <KUR>			
A;Cross-references: GB:AE007317; PIDN:AAK99115.1; PID:g15457866; GSPDB:GN00174			
C;Genetics:			
A;Gene: spr0311			
Query Match 72.2%; Score 13; DB 2; Length 57;			
Best Local Similarity 40.0%; Pred. No. 6.5e+02;			
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1 YXXXF 5		
Db	33 YSSTF 37		

Best Local Similarity 40.0%; Pred. No. 4.9e+02;		Mismatches 3; Indels 0; Gaps 0;	
Matches 2; Conservative 0;			
QY	1 YXXXF 5		
Db	37 YSAAF 41		
RESULT 6			
G82600			
hypothetical protein XF2098 [imported] - Xylella fastidiosa (strain 9a5c)			
C;Species: Xylella fastidiosa			
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000			
C;Accession: G82600			
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen			
Nature 406, 151-157, 2000			
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.			
A;Reference number: A82515; MUID:20365717; PMID:10910347			
A;Note: for a complete list of authors see reference number A59328 below			
A;Accession: G82600			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-68 <SIM>			
A;Cross-references: GB:AE004025; GB:AE003849; NID:g9107217; PIDN:AAF84897.1; GSPDB:GN001.			
A;Experimental source: strain 9a5c			
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; AJ			
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrez, H.			
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.			
submitted to GenBank, June 2000			
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme			
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre			
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.			
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.			
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.			
Rodrigues, V.; Rosa, A.C.R. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki			
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira			
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z			
A;Reference number: A59328			
A;Contents: annotation			
C;Genetics:			
A;Gene: XF0336			
Query Match 72.2%; Score 13; DB 2; Length 58;			
Best Local Similarity 40.0%; Pred. No. 6.6e+02;			
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1 YXXXF 5		
Db	37 YSAAF 41		

A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF2098

Query Match 72.2%; Score 13; DB 2; Length 68;  
Best Local Similarity 40.0%; Pred. No. 7.5e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXP 5  
|  
Db 28 YAAFP 32

## RESULT 7

177547  
hypothetical protein 2 - Escherichia coli insertion sequence IS903

C:Species: Escherichia coli  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 20-Sep-1999

C:Accession: I77547  
R:Molet, B.; Iida, S.; Arber, W.  
Mol. Gen. Genet. 199, 534-536, 1985  
A:Title: An active variant of the prokaryotic transposable element IS903 carries an amb  
A:Reference number: I57738; MUID:85295477; PMID:2993802  
A:Accession: I77547  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-70 <RES>  
A:Cross-references: EMBL:X02527; NID:g43694; PIDN:CAA26363.1; PID:g43696  
A:Experimental source: strain K-12, subspecies WA921

C:Genetics:  
C:Mobile element: insertion sequence IS903  
C:Superfamily: hypothetical protein IR903

Query Match 72.2%; Score 13; DB 2; Length 70;

Best Local Similarity 40.0%; Pred. No. 7.7e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXP 5  
|  
Db 14 YTTAF 18

## RESULT 8

G89851

hypothetical protein [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: G89851  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: G89851

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-72 <KUR>

A:Cross-references: GB:BA000018; PID:g13700673; PIDN:BA841970.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA0737

Query Match 72.2%; Score 13; DB 2; Length 72;

Best Local Similarity 40.0%; Pred. No. 7.8e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXP 5  
|  
Db 63 YTTSP 67

## RESULT 9

H97803

proline/betaine transporter RC0832 homolog [imported] - Rickettsia conorii (strain Malist  
C:Species: Rickettsia conorii

C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001

C:Accession: H97803

R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro

Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A:Reference number: A97700; MUID:21442074; PMID:11557893

A:Accession: H97803

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-75 <KUR>

A:Cross-references: GB:AB006914; PIDN:AAL03370.1; PID:g15619933; GSPDB:GN00173

C:Genetics:

A:Gene: RC0832

Query Match 72.2%; Score 13; DB 2; Length 75;

Best Local Similarity 40.0%; Pred. No. 8.1e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXP 5  
|  
Db 59 YSSAP 63

## RESULT 10

E49786

bacteriocin probable secretion protein A2 - Lactococcus lactis subsp. cremoris (strain 91  
C:Species: Lactococcus lactis subsp. cremoris

C:Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 20-Mar-1998

C:Accession: E49786

R:van Belkum, M.J.; Hayema, B.J.; Jeeninga, R.E.; Kok, J.; Venema, G.

Appl. Environ. Microbiol. 57, 492-498, 1991

A:Title: Organization and nucleotide sequences of two lactococcal bacteriocin operons.

A:Reference number: A49786; MUID:91197113; PMID:1901707

A:Accession: E49786

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-77 <VAN>

C:Genetics:

A:Genome: plasmid

Query Match 72.2%; Score 13; DB 2; Length 77;

Best Local Similarity 40.0%; Pred. No. 8.3e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXP 5  
|  
Db 12 YSSSF 16

## RESULT 11

AD1761

B. subtilis CsbA protein homolog lin2633 [imported] - Listeria innocua (strain Clip11262)  
C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AD1761

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Bloecker, ;

Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fslhi, H.;

Science 294, 849-852, 2001

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1761

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-79 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC97860.1; PID:g16415170; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:  
A:Gene: lin2633

Query Match 72.2%; Score 13; DB 2; Length 79;  
Best Local Similarity 40.0%; Pred. No. 8.5e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXP 5  
| |  
DB 44 YTSSP 48

## RESULT 12

JT0457  
hypothetical protein, 9K - fowlpox virus (isolate HP-438 [Munich])  
N:Alternate names: hypothetical protein c  
C:Species: fowlpox virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 21-Jul-2000  
C:Accession: JT0457  
R:Tomley, F.; Binns, M.; Campbell, J.; Boursnell, M.  
J. Gen. Virol. 69, 1025-1040, 1988  
A:Title: Sequence analysis of an 11.2 kilobase, near-terminal, BamHI fragment of fowlpox virus  
A:Reference number: JT0442; MUID:88229622; PMID:2836548  
A:Accession: JT0457  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-84 <TOM>  
A:Cross-references: GB:D00295; NID:g221380; PIDN:BAA00195.1; PID:g221386

Query Match 72.2%; Score 13; DB 2; Length 84;  
Best Local Similarity 40.0%; Pred. No. 8.9e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXP 5  
| |  
DB 30 YSSTF 34

## RESULT 13

T03993  
hypothetical protein TSL19.20 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: T03993  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15184  
A:Accession: T03993  
A:Molecule type: DNA  
A:Residues: 1-87 <BEV>  
A:Cross-references: EMBL:AL049481  
A:Experimental source: cultivar Columbia; BAC clone TSL19  
C:Genetics:  
A:Map position: 4  
A:Note: TSL19.20

Query Match 72.2%; Score 13; DB 2; Length 87;  
Best Local Similarity 40.0%; Pred. No. 9.2e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXP 5  
| |  
DB 11 YSSSF 15

## RESULT 14

C81917  
hypothetical protein NMA0737 [imported] - Neisseria meningitidis (strain Z2491 serogroup  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: C81917  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919

A:Accession: C81917  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-90 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84021.1; PID:g737945;  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0737

Query Match 72.2%; Score 13; DB 2; Length 90;  
Best Local Similarity 40.0%; Pred. No. 9.4e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXP 5  
| |  
DB 10 YAAAF 14

## RESULT 15

A81186  
hypothetical protein NMB0555 [imported] - Neisseria meningitidis (strain MC58 serogroup E  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: A81186  
R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: A81186  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-90 <TEST>  
A:Cross-references: GB:AE002411; GB:AE002098; NID:g7225776; PIDN:AAF40983.1; PID:g7225781  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0555

Query Match 72.2%; Score 13; DB 2; Length 90;  
Best Local Similarity 40.0%; Pred. No. 9.4e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXP 5  
| |  
DB 10 YTAAP 14

Search completed: October 5, 2004, 16:14:00  
Job time : 9.53425 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:49:36 ; Search time 4.89041 Seconds  
(without alignments)  
74.532 Million cell updates/sec

Title: US-09-973-473a-28

Perfect score: 18

Sequence: 1 YXXFPXX 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	72.2	38	PSAI_SYNEL	P25900 synechococ
2	13	72.2	46	PSAI_SYNP2	Q54752 synechococ
3	13	72.2	46	PSBK_PEA	P28642 pisum sativ
4	13	72.2	59	PSBK_SPIOL	P12163 spinacia ol
5	13	72.2	89	CFA_CITFR	P45509 citrobacter
6	13	72.2	93	YCXI_PAVLU	Q01572 pavlova lut
7	13	72.2	108	FLO9_YEAST	P39711 saccharomyc
8	13	72.2	108	FLO9_YEAST	Q10493 schizosacch
9	13	72.2	115	YD05_SCHPO	Q8X945 buchnera ap
10	13	72.2	119	Y531_BUCAL	P57597 buchnera ap
11	13	72.2	120	USH3_HUMAN	P58418 homo sapien
12	13	72.2	121	YK97_MYCTU	Q10696 mycobacteri
13	13	72.2	121	YK97_MYCTU	P34401 caenorhabdi
14	13	72.2	148	CADF_DROME	P45594 drosophila
15	13	72.2	149	Y38A_MYCTG	Q92b71 mycoplasma
16	13	72.2	154	Y1E7_MYCTU	P71686 mycobacteri
17	13	72.2	154	YL66_ARCFU	O28116 archaeoglob
18	13	72.2	157	Y012_BPL2	P42547 bacterioph
19	13	72.2	161	YZCX_ECOLI	P11291 escherichia
20	13	72.2	162	KCH2_CAVPO	O08703 cavia porce
21	13	72.2	163	YCSL_BACUN	P30906 bacteroides
22	13	72.2	166	ING_SHEEP	P79154 capra hircu
23	13	72.2	166	ING_SHEEP	P17773 ovis aries
24	13	72.2	167	VHR2_VABAM	Q9qbb4 yaba monkey
25	13	72.2	175	LICT_ECOLI	P37025 escherichia
26	13	72.2	178	DUSP_MXXVL	Q85297 myxoma viru
27	13	72.2	178	VHR2_YLDV	Q9dhp6 yaba-like d
28	13	72.2	181	RL5_METVA	P14029 methanococc
29	13	72.2	183	SFP4_BOVIN	P81019 bos taurus
30	13	72.2	193	NU2M_PARTE	P15577 paramecium
31	13	72.2	194	PUP0_HALANI	Q9bm4 halobacteri
32	13	72.2	205	HEW2_CLOOQ	Q59295 clostridium
33	13	72.2	206	YD94_MYCPN	P75483 mycoplasma

34	13	72.2	208	1	XLR_MOUSE	P05531 mus musculu
35	13	72.2	216	1	NRGD_HUMAN	P26718 homo sapien
36	13	72.2	216	1	NRGD_MACMU	Q9mzj7 macaca muli
37	13	72.2	219	1	EUTL_ECOLI	P76541 escherichia
38	13	72.2	219	1	EUTL_SALTY	Q9zfu9 salmonella
39	13	72.2	223	1	TPIS_AERPE	Q9ybr1 aeropyrum p
40	13	72.2	227	1	CIS3_YEAST	P47001 saccharomyc
41	13	72.2	227	1	RHON_HUMAN	P52198 homo sapien
42	13	72.2	227	1	RHON_MOUSE	Q9qym5 mus musculu
43	13	72.2	228	1	GLUC_COREF	Q8tql5 corynebacte
44	13	72.2	228	1	GLUC_CORGL	P48244 corynebacte
45	13	72.2	228	1	MX11_HUMAN	P50539 homo sapien

#### ALIGNMENTS

##### RESULT 1

PSAI\_SYNEL STANDARD; PRT; 38 AA.  
AC P25900;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Photosystem I reaction center subunit VIII.  
GN PSAI OR TSR2405.  
OS Synechococcus elongatus (Thermosynechococcus elongatus), and  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=32046, 1141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.elongatus; STRAIN=BP-1;  
RX MEDLINE=22225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium  
Thermosynechococcus elongatus BP-1.";  
RL DNA Res. 9:123-130(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.e.naegeli;  
RX MEDLINE=97057537; PubMed=84862290;  
RA Muehlenhoff U., Haehnel W., Witt H.T., Herrmann R.G.;  
RT "Genes encoding eleven subunits of photosystem I from the  
thermophilic cyanobacterium Synechococcus sp.";  
RL Gene 127:71-78(1993).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (4.0 ÅNGSTROMS).  
RC SPECIES=S.e.naegeli;  
RX MEDLINE=97057537; PubMed=8901876;  
RA Krauss N., Schubert W.D., Klukas O., Fromme P., Witt H.T., Saenger W.;  
RT "Photosystem I at 4-Å resolution represents the first structural  
model of a joint photosynthetic reaction centre and core antenna  
system.";  
RL Nat. Struct. Biol. 3:965-973(1996).  
CC -!- FUNCTION: May help in the organization of the psal subunit.  
CC -!- SIMILARITY: Belongs to the psal family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
EMBL; AP005377; BAC09957.1; -  
EMBL; X53763; CAA45297.1; -  
PDB; 2PPS; 27-MAY-98.  
HAMAP; MF\_00431; -; 1.

```
DR InterPro; IPR001302; PSI_8.
DR Pfam; PF00796; PSI_8; 1.
DR ProDom; PD003995; PSI_8; 1.
KW Photosystem I; Photosynthesis; Transmembrane; 3D-structure;
KW Complete proteome. 12 32 POTENTIAL.
FT TRANSMEM 12 32
SQ SEQUENCE 38 AA; 4297 MW; E9B0178560D85CF5 CRC64;

Query Match 72.2%; Score 13; DB 1; Length 38;
Best Local Similarity 40.0%; Pred. No. 2.8e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXX 5
| |
Db 5 YAAAF 9

RESULT 2
PSAI_SYNP2
ID -PSAI_SYNP2 STANDARD; PRT; 38 AA.
AC Q54752;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem I reaction center subunit VIII.
GN PSAI.
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32049;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96271772; PubMed=8787020;
RA Schlachter W.M., Shen G., Zhao J., Bryant D.A.;
RT "Characterization of psal and psal mutants of Synechococcus sp.
RT strain PCC 7002: a new model for state transitions in
RT cyanobacteria.";
RL Photochem. Photobiol. 64:53-66(1996).
CC -!- FUNCTION: May help in the organization of the psal subunit.
CC -!- SIMILARITY: Belongs to the psal family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U58035; AAB18909.1; -.
CC HAMAP; MF 00431; -.
CC InterPro; IPR001302; PSI_8.
CC Pfam; PF00796; PSI_8; 1.
KW Photosystem I; Photosynthesis; Transmembrane.
FT TRANSMEM 9 29
SQ SEQUENCE 38 AA; 3957 MW; BE58C30EFAB31832 CRC64;

Query Match 72.2%; Score 13; DB 1; Length 38;
Best Local Similarity 40.0%; Pred. No. 2.8e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXX 5
| |
Db 5 YAAAF 9

RESULT 3
PSBK_PEA
ID -PSBK_PEA STANDARD; PRT; 46 AA.
AC P28642; Q9T2J7;
DT 01-DEC-1992 (Rel. 24, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem II reaction center protein K precursor (PSII-K).
GN Photosystem II reaction center protein K precursor (PSII-K).
OS Spinacia oleracea (Spinach).

DE (Fragment).
GN PSBK.
OS Pisum sativum (Garden pea).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE OF 1-29 FROM N.A.
RC STRAIN=cv. Alaska;
RX MEDLINE=92224289; PubMed=1807835;
RA Nagano Y., Matsuno R., Sasaki Y.;
RT "Sequence and transcriptional analysis of the gene cluster trnQ-zfpA-
RT psal-ORF231-petA in pea chloroplasts.";
RL Curr. Genet. 20:431-436(1991).
RN [2]
RP SEQUENCE OF 25-32, AND MASS SPECTROMETRY.
RX MEDLINE=98298118; PubMed=9632665;
RA Zheleva D., Sharma J., Panico M., Morris H.R., Barber J.;
RT "Isolation and characterization of monomeric and dimeric CP47-reaction
RT center photosystem II complexes.";
RL J. Biol. Chem. 273:16122-16127(1998).
RN [3]
RP SEQUENCE OF 25-46.
RX MEDLINE=94085601; PubMed=8262226;
RA Zakharov S.D., Ewy R.G., Dilley R.A.;
RT "Subunit III of the chloroplast ATP-synthase can form a Ca(2+)-binding
RT site on the luminal side of the thylakoid membrane.";
RL FEBS Lett. 336:95-99(1993).
CC -!- FUNCTION: This protein is a component of the reaction center of
CC photosystem II.
CC -!- SIMILARITY: Belongs to the psbK family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X56315; CAA39753.1; -.
CC PIR; S17919; S17919.
CC HAMAP; MF 00441; -.
CC InterPro; IPR003687; PSII_PsbK.
CC Pfam; PF02533; PsbK; 1.
KW Photosystem II; Chloroplast.
FT PROPEP 1 24
FT CHAIN 25 >46
FT CONFLICT 31 31
FT NON_TER 46 46
SQ SEQUENCE 46 AA; 5126 MW; 382B64B0C5417633 CRC64;

Query Match 72.2%; Score 13; DB 1; Length 46;
Best Local Similarity 40.0%; Pred. No. 3.3e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXX 5
| |
Db 17 YSSSF 21

RESULT 4
PSBK_SPIOL
ID -PSBK_SPIOL STANDARD; PRT; 59 AA.
AC P12163; Q9M3M8;
DT 01-OCT-1989 (Rel. 12, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem II reaction center protein K precursor (PSII-K).
GN PSBK.
OS Spinacia oleracea (Spinach).
```

OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllales; Amaranthaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.  
RA Murata N., Miyao M., Hayashida N., Hidaka T., Sugiura M.;  
RT "Identification of a new gene in the chloroplast genome encoding a  
RL low-molecular-mass polypeptide of photosystem II complex.";  
RN FEBS Lett. 235:283-288(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=sv. Geant d'hiver, and cv. Monatol;  
RX MEDLINE=21187424; PubMed=11292076;  
RA Schmitz-Linneweber C., Maier R.M., Alcaraz J.-P., Cottet A.,  
RA Hermann R.G., Mache R.;  
RT "The plastid chromosome of spinach (Spinacia oleracea): complete  
RL nucleotide sequence and gene organization.";  
RN Plant Mol. Biol. 45:307-315(2001).  
RN [3]  
RP SEQUENCE OF 23-34.  
RA Schroeder W.P., Henrysson T., Akerlund H.E.;  
RT "Characterization of low molecular mass proteins of photosystem II by  
RL N-terminal sequencing.";  
RN FEBS Lett. 235:289-292(1988).  
RN [4]  
RP SEQUENCE OF 23-36.  
RX MEDLINE=89121082; PubMed=2644131;  
RA Ikeuchi M., Takio K., Inoue Y.;  
RT "N-terminal sequencing of photosystem II low-molecular-mass proteins.  
RL 5 and 4.1 kDa components of the O2-evolving core complex from higher  
RN plants.";  
RN FEBS Lett. 242:263-269(1989).  
RN [5]  
RP SEQUENCE OF 23-30, AND MASS SPECTROMETRY.  
RX MEDLINE=98298118; PubMed=9632665;  
RA Zhelleva D., Sharma J., Panico M., Morris H.R., Barber J.;  
RT "Isolation and characterization of monomeric and dimeric CP47-reaction  
RL center photosystem II complexes.";  
RN J. Biol. Chem. 273:16122-16127(1998).  
CC -!- FUNCTION: This protein is a component of the reaction center of  
CC photosystem II.  
CC -!- MASS SPECTROMETRY: MW=4292.1; METHOD=MALDI; RANGE=23-59.  
CC -!- SIMILARITY: Belongs to the psbK family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X12786; CAA31277.2; ALT INIT.  
DR EMBL; AJ400848; CAB88708.1; --  
DR HAMAP; MF\_00441; -- 1.  
DR InterPro; IPR003687; PSII\_PsbK.  
DR Pfam; PF02533; PsbK; 1.  
KW Photosystem II; Chloroplast.  
FT PROPEP 1 22  
FT CHAIN 23 59  
FT PHOTOSYSTEM II REACTION CENTER PROTEIN K.  
FT I -> T (IN REF. 1).  
FT I -> IGI (IN REF. 1).  
FT ALY -> TUF (IN REF. 1).  
FT CONFLICT 8 8  
FT CONFLICT 13 15  
FT CONFLICT 32 32  
FT CONFLICT 37 37  
FT CONFLICT F -> I (IN REF. 1).  
SQ SEQUENCE 59 AA; 6749 MW; 25FCFA8925CE157F CRC64;  
Query Match 72.2%; Score 13; DB 1; Length 59;  
Best Local Similarity 40.0%; Pred. No. 4e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
-----  
OY 1 YXXXF 5  
DB 15 YSSSF 19  
-----  
RESULT 5  
CFA\_CITFR STANDARD; PRT; 89 AA.  
AC P45509;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Cyclopropane-fatty-acyl-phospholipid synthase (EC 2.1.1.79)  
DE (Cyclopropane fatty acid synthase) (CFA synthase) (Fragment).  
GN CFA.  
OS Citrobacter freundii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Citrobacter.  
OX NCBI\_TaxID=546;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 30040;  
RA Daniel R., Gottschalk G.;  
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: TRANSFERS A METHYLENE GROUP FROM S-ADENOSYL-L-METHIONINE  
CC TO THE CIS DOUBLE BOND OF AN UNSATURATED FATTY ACID CHAIN  
CC RESULTING IN THE REPLACEMENT OF THE DOUBLE BOND WITH A METHYLENE  
CC BRIDGE (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + phospholipid  
CC olefinic fatty acid = S-adenosyl-L-homocysteine + phospholipid  
CC cyclopropane fatty acid.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U09771; AAB4842.1; --  
DR InterPro; IPR003333; CWAS.  
DR Pfam; PF02353; CWAS; 1.  
KW Transferase; Methyltransferase; Lipid synthesis.  
FT NON TER 1  
SQ SEQUENCE 89 AA; 10609 MW; 7D0DD89F377F72EA CRC64;  
Query Match 72.2%; Score 13; DB 1; Length 89;  
Best Local Similarity 40.0%; Pred. No. 5.5e+02;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
-----  
OY 1 YXXXF 5  
DB 46 YSATF 50  
-----  
RESULT 6  
YCX1\_PAVLUJ STANDARD; PRT; 93 AA.  
AC Q01572;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 10.9 kDa protein in SECA 5' region.  
OS Pavlova lutheri (Monochrysis lutheri).  
OG Chloroplast.  
OC Eukaryota; Haptophyceae; Pavlova.  
OX NCBI\_TaxID=2832;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93046814; PubMed=1423730;  
RA Scaramuzzi C.D., Hiller R.G., Stokes H.W.;

RT "Identification of a chloroplast-encoded *secA* gene homologue in a  
 RT chromophytic alga: possible role in chloroplast protein  
 RL translocation.";  
 RL Curr. Genet. 22:421-427(1992).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X65961; CAA46775.1; -.  
 DR PIR; S27028; S27028.  
 KW Chloroplast; Hypothetical protein.  
 SQ SEQUENCE 93 AA; 10860 MW; 3F5D17E16B03A4F CRC64;  
 Query Match 72.2%; Score 13; DB 1; Length 93;  
 Best Local Similarity 40.0%; Pred. No. 5.7e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 YXXXP 5  
 Db 57 YTTSP 61  
 -----  
 RESULT 7  
 FLO9\_YEAST STANDARD; PRT; 108 AA.  
 AC P39711;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Flocculation protein FLO9.  
 GN FLO9 OR YAL064W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.  
 OX NCBI\_TaxID=4932;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=95249563; PubMed=7731988;  
 RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,  
 RA Hall J., Ouellette B.F.P., Keng T., Barton A.B., Su Y., Davies C.K.,  
 RA Storms R.K.;  
 RA "The nucleotide sequence of chromosome I from Saccharomycos  
 RT cerevisiae";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).  
 CC -!- SIMILARITY: Belongs to the flocculin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U12980; AAC04970.1; -.  
 DR PIR; S51958; S51958.  
 DR GeneOnline; 138402; -.  
 DR SGD; S0000059; FLO9.  
 SQ SEQUENCE 108 AA; 12772 MW; 2BF3D67501A7E3D2 CRC64;  
 Query Match 72.2%; Score 13; DB 1; Length 108;  
 Best Local Similarity 40.0%; Pred. No. 6.4e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 YXXXP 5  
 Db 3 YTAIF 7

RESULT 8  
 YDGS\_SCHPO STANDARD; PRT; 115 AA.  
 ID YDGS\_SCHPO  
 AC Q10493;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE Very hypothetical protein C26F1.05 in chromosome I.  
 DE SPAC26F1.05.  
 GN Schizosaccharomyces pombe (Fission yeast).  
 OS Schizosaccharomycetes  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OX NCBI\_TaxID=4896;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Welljens I., Vansteens E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomycos pombe";  
 RL Nature 415:871-880(2002).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z73100; CAA97362.1; -.  
 DR PIR; T38415; T38415.  
 DR GeneDB Spombe; SPAC26F1.05; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 115 AA; 13477 MW; B0C842741F30326D CRC64;  
 Query Match 72.2%; Score 13; DB 1; Length 115;  
 Best Local Similarity 40.0%; Pred. No. 6.7e+02;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 YXXXP 5  
 Db 103 YTTST 107  
 -----  
 RESULT 9  
 Y512\_BUCAP STANDARD; PRT; 119 AA.  
 ID Y512\_BUCAP

```

AC O8K945;
DT 28-FEB-2003 (Rel. 41, Created)
DE Hypothetical protein BUE512.
GN BUE512.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamás I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC -!- SIMILARITY: Belongs to the UPF0116 (dsrF) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE014125; AAM68055.1; -.
DR HAMAP; MF 00389; -. 1.
DR InterPro; IPR003787; DsrE.
DR Pfam; PF02635; DsrE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 119 AA; 13610 MW; 6E924AD44058CC0 CRC64;

Query Match 72.2%; Score 13; DB 1; Length 119;
Best Local Similarity 40.0%; Pred. No. 6.9e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXP 5
DB 61 YTSSP 65

RESULT 10
Y531_BUCAI
ID Y531_BUCAI STANDARD; PRT; 119 AA.
AC P57597;
DT 16-OCT-2001 (Rel. 40, Created)
DE Hypothetical protein BUE531.
GN BUE531.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- SIMILARITY: Belongs to the UPF0116 (dsrF) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF388366; AAL09581.1; -.
DR EMBL; AF388368; AAL09582.1; -.
DR Genbank; HGNC:12605; USH3A.
DR MIM; 606397; -.

or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001119; BAB13224.1; -.
CC HAMAP; MF 00389; -. 1.
CC InterPro; IPR003787; DsrE.
CC Pfam; PF02635; DsrE; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 1119 AA; 13506 MW; CS95F1215BE10938 CRC64;

Query Match 72.2%; Score 13; DB 1; Length 119;
Best Local Similarity 40.0%; Pred. No. 6.9e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXP 5
DB 61 YTSSP 65

RESULT 11
USH3_HUMAN
ID USH3_HUMAN STANDARD; PRT; 120 AA.
AC P58418;
DT 28-FEB-2003 (Rel. 41, Created)
DE Hypothetical protein; Complete proteome.
DT 10-OCT-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Usher syndrome type 3 protein.
GN USH3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND B), AND VARIANTS USH3 LYS-44 AND
RP 77-ILE-LEU-78 DELINS MET.
RX MEDLINE=21426338; PubMed=11524702;
RA Joensuu T., Haemaelsinen R., Yuan B., Johnson C., Tegelberg S.,
RA Gasparini P., Zelance J., Pirvola U., Pakarinen L., Lehesjoki A.-E.,
RA de la Chapelle A., Sankila E.-M.;
RT "Mutations in a novel gene with transmembrane domains underlie Usher
RT syndrome type 3.";
RL Am. J. Hum. Genet. 69:673-684(2001).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=A;
CC IsoId=P58418-1; Sequence=Displayed;
CC Name=B;
CC IsoId=P58418-2; Sequence=VSP_004008, VSP_004009;
CC -!- TISSUE SPECIFICITY: Widely expressed. Found in the retina.
CC -!- DISEASE: Defects in USH3A are the cause of Usher syndrome type 3
CC (USH3) [MIM:276902]. Usher syndrome is an autosomal recessive
CC condition, characterized by the association of retinitis
CC pigmentosa with sensorineural deafness. Patients with USH3 have
CC progressive hearing loss, variably present vestibular dysfunction
CC and adult onset retinitis pigmentosa. USH3 is a common form of
CC Usher syndrome in Finland, where it accounts for 42% of all Usher
CC syndrome cases.
CC -!- DATABASE: NAME=Mutations of the USH3A gene;
CC NOTE=Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/ush3mut.htm".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF388366; AAL09581.1; -.
CC EMBL; AF388368; AAL09582.1; -.
CC Genbank; HGNC:12605; USH3A.
CC MIM; 606397; -.

```

```

DR MIM; 276902; -.
KW Transmembrane; Alternative splicing; Vision; Disease mutation;
KW Deafness; Retinitis pigmentosa; Usher syndrome.
FT DOMAIN 1 24 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 46 58 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 80 120 CYTOPLASMIC (POTENTIAL).
FT VARSPLIC 9 30 VPDLKAIPIVSHVNVILFSA -> GYKLCTTGILSLIF
FT YGTSNT (in isoform B).
FT VARSPLIC 31 120 /FTID=VSP 004008.
FT VARIANT 44 44 Missing (in isoform B).
FT VARIANT 44 44 /FTID=VSP 004009.
FT VARIANT 77 78 M -> K (in USH3).
FT VARIANT 77 78 /FTID=VAR 012241.
FT VARIANT 77 78 IL -> M (in USH3).
FT VARIANT 77 78 /FTID=VAR 012242.
SQ SEQUENCE 120 AA; 13421 MW; 959B081E7665A2D1 CRC64;

Query Match 72.2%; Score 13; DB 1; Length 120;
Best Local Similarity 40.0%; Pred. No. 7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 109 YTTSF 113

RESULT 12
YK87_MYCTU
ID YK87_MYCTU STANDARD; PRT; 121 AA.
AC Q10696;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein Rv2087/MT2148.
GN Rv2087 OR MT2148 OR MFCY49.27.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.P., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -1- SIMILARITY: SOME, TO TRANSPOSASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; 273966; CAA98199.1; -.
DR EMBL; AE007064; -; NOT_ANNOTATED_CDS.
DR PIR; B70767; B70767.
DR TIGR; MT2148; -.
DR YGTSNT; Rv2087; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 121 AA; 12996 MW; 020E92098EFFC0C4 CRC64;

Query Match 72.2%; Score 13; DB 1; Length 121;
Best Local Similarity 40.0%; Pred. No. 7e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 11 YATAF 15

RESULT 13
YLU7_CAEEL
ID YLU7_CAEEL STANDARD; PRT; 131 AA.
AC P34401;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein F10E9.7 in chromosome III.
GN F10E9.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton J., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smallon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Whidman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RN Nature 368:32-38(1994).
RP REVISIONS.
RA Waterston R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=22206494; PubMed=12218036;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.P., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -1- SIMILARITY: SOME, TO TRANSPOSASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; L10986; AAA28017.3; -.
DR PIR; S44805; S44805.
DR WormPep; F10E9.7; CE29494.
DR InterPro; IPR002125; dCWP/cyt_deam.
DR Pfam; PF00383; dCWP_cyt_deam; 1.
KW Hypothetical protein.
SQ SEQUENCE 131 AA; 14888 MW; F22AF23E54C6C315 CRC64;

```

```

Query Match      72.2%; Score 13; DB 1; Length 131;
Best Local Similarity 40.0%; Pred. No. 7.5e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 YXXXP 5
DB 69 YTSFP 73

RESULT 14
CADEF DROME STANDARD; PRT; 148 AA.
AC P45594; Q9W1C4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cofilin/actin depolymerizing factor homolog (D61 protein) (Twinstar protein).
GN TSR OR CADP OR CG4254.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94240181; PubMed=8183953;
RA Edwards K.A., Montague R.A., Shepard S., Edgar B.A., Erikson R.L.,
RA Kiehart D.P.;
RT Identification of Drosophila cytoskeletal proteins by induction of
RT abnormal cell shape in fission yeast.
RL Proc. Natl. Acad. Sci. U.S.A. 91:4589-4593(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=96095784; PubMed=8522587;
RA Gunsalus K.C., Bonaccorsi S., Williams E., Verni F., Gatti M.,
RA Goldberg M.L.;
RT "Mutations in twinstar, a Drosophila gene encoding a cofilin/ADF
RT homologue, result in defects in centrosome migration and
RT cytokinesis."
RL J. Cell Biol. 131:1243-1259(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.E., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hovsin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

```

```

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector A.C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Zhao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Belongs to the actin-binding proteins ADF family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U08217; AAA19856.1; -.
DR EMBL: U24490; AAC46962.1; -.
DR EMBL: U24676; AAC46963.1; -.
DR EMBL: AE003462; AAF47146.1; -.
DR PIR: A57569; A57569.
DR HSP: Q39250; 1F7S.
DR FlyBase; FBgn0011726; tsr.
DR GO: GO:0003779; F:actin binding; IMP.
DR InterPro; IPR002108; Actbind_cofin.
DR Pfam; PF00241; cofilin_ADF; 1.
DR ProDom; PD002129; Actbind_cofin; 1.
DR SMART; SM00102; ADF; 1.
DR PROSITE; PS00325; ACTIN DEPOLYMERIZING; 1.
KW Actin-binding; Cytoskeleton; Nuclear protein.
FT DOMAIN 19 23 ACTIN LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 96 115 ACTIN-BINDING (POTENTIAL).
SQ SEQUENCE 148 AA; 17153 MW; 24F7216033859620 CRC64;

Query Match      72.2%; Score 13; DB 1; Length 148;
Best Local Similarity 40.0%; Pred. No. 8.2e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 YXXXP 5
DB 107 YSSSF 111

RESULT 15
Y38A MYCGE STANDARD; PRT; 149 AA.
AC Q92B71;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG384.1.
GN MG384.1.
CN Mycoplasma genitalium.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."

```

```
RL Science 270:397-403(1995).
RN [2]
RP IDENTIFICATION.
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
RL -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39720; AAC71616.1; -.
DR TIGR; MG384.1; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
SQ SEQUENCE 149 AA; 17731 MW; F0A7A8DD2562384 CRC64;
Query Match 72.2%; Score 13; DB 1; Length 149;
Best Local Similarity 40.0%; Pred. NO. 8.3e+02;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 YXXXP 5
Db 47 YSTAF 51
```

Search completed: October 5, 2004, 16:07:34  
Job time : 4.89041 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2004, 15:56:41 ; Search time 27.2329 Seconds  
(without alignments)  
81.102 Million cell updates/sec

Title: US-09-973-473a-28

Perfect score: 18

Sequence: 1 YXXFXFX 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mmc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	72.2	32	Q35494	Q35494 paracentrot
2	13	72.2	39	Q87H11	Q87H11 vibrio para
3	13	72.2	40	Q9JNG3	Q9JNG3 streptococc
4	13	72.2	40	Q97S64	Q97S64 streptococc
5	13	72.2	42	Q8KDF7	Q8KDF7 chlorobium
6	13	72.2	42	Q8FHX5	Q8FHX5 escherichia
7	13	72.2	45	Q9DFX1	Q9DFX1 gilllichthys
8	13	72.2	45	Q8KDN9	Q8KDN9 chlorobium
9	13	72.2	49	Q8W6Q5	Q8W6Q5 bacterioph
10	13	72.2	56	Q8EW08	Q8EW08 mycoplasma
11	13	72.2	57	Q9U7H8	Q9U7H8 plasmodium
12	13	72.2	57	Q8CZ69	Q8CZ69 streptococc
13	13	72.2	58	Q9PGG6	Q9PGG6 xylella fas
14	13	72.2	59	Q8GRV3	Q8GRV3 oryza sativ
15	13	72.2	60	Q9DGS1	Q9DGS1 amsacta moo
16	13	72.2	60	Q82PY6	Q82PY6 streptomyce

## ALIGNMENTS

### RESULT 1

Q35494 PRELIMINARY; PRT; 32 AA.  
AC Q35494;  
DT 01-NOV-1996 (TremBLrel. 01, Created)  
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Cytochrome oxidase subunit 3 (EC 1.9.3.1) (Cytochrome c oxidase  
DE polypeptide III) (Fragment).  
GN COIII.  
OS Paracentrotus lividus (Common sea urchin).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoida; Euechinoida; Echinacea; Echinoida; Echinidae;  
OC Paracentrotus.  
OX NCBI\_TaxID=7656;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87248108; PubMed=3596250;  
RA Cantatore P., Roberti M., Morisco P., Rainaldi G., Gadaleta M.N.,  
RA Saccone C.;  
RT "A novel gene order in the Paracentrotus lividus mitochondrial  
RT genome.";  
RL Gene 53:41-54(1987).  
CC -!- FUNCTION: SUBUNITS I, II AND III FORM THE FUNCTIONAL CORE OF THE  
CC ENZYME COMPLEX (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
CC C + 2 H(2)O.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 3 FAMILY.  
DR EMBL; M16524; AAA31995.2; -;  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0005739; C:mitochondrion; IEA.  
DR GO; GO:0009481; F:aa3-type cytochrome c oxidase; IEA.  
DR GO; GO:0009482; F:ba3-type cytochrome c oxidase; IEA.  
DR GO; GO:0009483; F:caa3-type cytochrome c oxidase; IEA.  
DR GO; GO:0009485; F:cb3-type cytochrome c oxidase; IEA.  
DR GO; GO:0004129; F:cytochrome-c oxidase activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.

17 13 72.2 61 5 Q95ZM3  
18 13 72.2 61 10 Q8S8Y4  
19 13 72.2 61 12 Q8VIB6  
20 13 72.2 68 16 Q9PBP2  
21 13 72.2 68 16 Q87DA7  
22 13 72.2 69 2 Q8RIH9  
23 13 72.2 69 16 Q8FUR6  
24 13 72.2 69 16 Q827X0  
25 13 72.2 70 2 Q48347  
26 13 72.2 72 16 Q99VJ9  
27 13 72.2 72 16 Q932C6  
28 13 72.2 73 5 Q9ND38  
29 13 72.2 73 10 Q9FF49  
30 13 72.2 74 12 Q9YPP8  
31 13 72.2 74 12 Q81696  
32 13 72.2 74 12 Q9YPP9  
33 13 72.2 74 13 Q8JFV7  
34 13 72.2 75 9 Q38481  
35 13 72.2 75 16 Q92HD9  
36 13 72.2 75 16 Q8XHA2  
37 13 72.2 76 12 Q68554  
38 13 72.2 79 16 Q928A3  
39 13 72.2 81 5 Q8T3U0  
40 13 72.2 81 17 Q8TLN7  
41 13 72.2 81 17 Q8Q069  
42 13 72.2 82 2 Q9X5W4  
43 13 72.2 82 16 Q82UE3  
44 13 72.2 84 12 Q9YFJ9  
45 13 72.2 84 12 Q80DK4

Q95zm3 caenorhabdi  
Q88y4 atropa bell  
Q8vib6 hepatitis c  
Q9pbp2 xylella fas  
Q87da7 xylella fas  
Q9rih9 streptococ  
Q8fjr6 escherichia  
Q827x0 streptomyce  
Q48347 escherichia  
Q99vj9 staphylococ  
Q932c6 staphylococ  
Q9nd38 plasmodium  
Q9fp49 oryza sativ  
Q9ypp8 hepatitis c  
Q81696 hepatitis c  
Q9ypp9 hepatitis c  
Q8jfv7 brachydanio  
Q38481 bacterioph  
Q92hd9 rickettsia  
Q8xha2 clostridium  
Q68554 hepatitis c  
Q928a3 listeria in  
Q8t3u0 drosophila  
Q8tln7 methanosarc  
Q9q069 methanosarc  
Q9x5w4 rhodobacter  
Q82ue3 nitrosomona  
Q9ypj9 fowlpox vir  
Q80dk4 hepatitis c

DR InterPro; IPR000298; CytC\_oxdse\_III.  
 DR Pfam; PF00510; COX3; 1.  
 DR ProDom; PD000382; CytC\_oxdse\_III; 1.  
 KW Oxidoreductase; TransmembranE; Mitochondrion.  
 FT NON\_TER 1  
 FT NON\_TER 32 32  
 SQ SEQUENCE 32 AA; 3793 MW; 6C298A6B440D6A35 CRC64;

Query Match 72.2%; Score 13; DB 8; Length 32;  
 Best Local Similarity 40.0%; Pred. No. 1.8e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
 | | |  
 Db 25 YSTTF 29

## RESULT 2

ID Q87H11 PRELIMINARY; PRT; 39 AA.  
 AC Q87H11;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN VPAl154.  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=670;  
 [1]  
 SEQUENCE FROM N.A.

RP STRAIN=RIMD 2210633 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;  
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
 distinct from that of V. cholerae.";  
 RL Lancet 361:743-749(2003).  
 DR EMBL; AP005088; BAC62497.1; --  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 39 AA; 4817 MW; EA265E7CD2141419 CRC64;

Query Match 72.2%; Score 13; DB 16; Length 39;  
 Best Local Similarity 40.0%; Pred. No. 2.1e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
 | | |  
 Db 23 YASTF 27

## RESULT 3

ID Q9JNG3 PRELIMINARY; PRT; 40 AA.  
 AC Q9JNG3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Sic1.232.  
 GN SITC.  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 [1]  
 SEQUENCE FROM N.A.

RP STRAIN=MGAS6644;  
 RX MEDLINE=98060444; PubMed=9399523;  
 RA Mejia L.M., Stockbauer K.E., Pan X., Cravioto A., Musser J.M.;  
 RT "Characterization of group A Streptococcus strains recovered from  
 Mexican children with pharyngitis by automated DNA sequencing of

RT virulence-related genes: unexpectedly large variation in the gene  
 RL (sic) encoding a complement-inhibiting protein.";  
 J. Clin. Microbiol. 35:3220-3224(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS6644;  
 RX MEDLINE=98169508; PubMed=9501227;  
 RA Stockbauer K.E., Grigsby D., Pan X., Fu Y.X., Mejia L.M., Cravioto A.,  
 RA Musser J.M.;  
 RT "Hypervariability generated by natural selection in an extracellular  
 complement-inhibiting protein of serotype M1 strains of group A  
 Streptococcus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:3128-3133(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS6644;  
 RX MEDLINE=99238853; PubMed=10221878;  
 RA Hoe N., Nakashima K., Grigsby D., Pan X., Dou S.J., Naidich S.,  
 RA Garcia M., Kahn E., Bergmire-Sweat D., Musser J.M.;  
 RT "Rapid molecular genetic subtyping of serotype M1 group A  
 Streptococcus strains.";  
 RL Emerging Infect. Dis. 5:254-263(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS6644;  
 RX MEDLINE=99353360; PubMed=10426317;  
 RA Hoe N.P., Nakashima K., Lukomski S., Grigsby D., Liu M., Kordari P.,  
 RA Dou S.-J., Pan X., Vuopio-Varkila J., Salmelinn S., McGee A.,  
 RA Low D.E., Schwartz B., Schuchat A., Naidich S., De Lorenzo D.,  
 RA Fu Y.-X., Musser J.M.;  
 RT "Rapid selection of complement-inhibiting protein variants in group A  
 Streptococcus epidemic waves.";  
 RL Nat. Med. 5:924-929(1999).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS6644;  
 RX PubMed=11015234;  
 RA Hoe N.P., Kordari P., Cole R., Liu M., Palzkill T., Huang W.,  
 RA McLellan D., Adams G.J., Hu M., Vuopio-Varkila J., Cate T.R.,  
 RA Pichichero M.E., Edwards K.M., Eskola J., Low D.E., Musser J.M.;  
 RT "Human immune response to streptococcal inhibitor of complement, a  
 serotype M1 group A Streptococcus extracellular protein involved in  
 epidemics.";  
 RL J. Infect. Dis. 182:1425-1436(2000).  
 DR EMBL; AP232537; AAF65001.1; --  
 DR InterPro; IPR005328; Sic.  
 DR Pfam; PF03482; sic; 1.  
 SQ SEQUENCE 40 AA; 4738 MW; 4C250CA3832D06CD CRC64;

Query Match 72.2%; Score 13; DB 2; Length 40;  
 Best Local Similarity 40.0%; Pred. No. 2.2e+03;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
 | | |  
 Db 27 YTTTF 31

## RESULT 4

ID Q97S64 PRELIMINARY; PRT; 40 AA.  
 AC Q97S64;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein SP0548.  
 GN SP0548.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 [1]  
 SEQUENCE FROM N.A.

```

RC STRAIN-ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.P., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007365; AAK4705.1; -.
DR PIR; H95063; H95063.
DR TIGR; SP0548; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 40 AA; 4563 MW; 52F75CA2P36FF187 CRC64;

Query Match 72.2%; Score 13; DB 16; Length 40;
Best Local Similarity 40.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXF 5
Db 29 YSTSF 33

RESULT 5
Q8KDF7 PRELIMINARY; PRT; 42 AA.
ID Q8KDF7
AC Q8KDF7
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein CT1093.
GN CT1093.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OC NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., DeBoy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR EMBL; AE012872; AAM72326.1; -.
DR TIGR; CT1093; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 42 AA; 4874 MW; 847FA4B0F406CCF1 CRC64;

Query Match 72.2%; Score 13; DB 16; Length 42;
Best Local Similarity 40.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXF 5
Db 29 YSSSF 33

RESULT 6
Q8FHX5 PRELIMINARY; PRT; 42 AA.
ID Q8FHX5
AC Q8FHX5

```

```

DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN C1702.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6.H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raako D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016760; AAN80169.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 42 AA; 4739 MW; F0571A4C8EE18B01 CRC64;

Query Match 72.2%; Score 13; DB 16; Length 42;
Best Local Similarity 40.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXF 5
Db 20 YASF 24

RESULT 7
Q9DFK1 PRELIMINARY; PRT; 45 AA.
ID Q9DFK1
AC Q9DFK1
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE C-4 sterol methyl oxidase (Fragment).
OS Gillichthys seta (Shortjaw mudsucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidi;
OC Gobiidae; Gillichthys.
OX NCBI_TaxID=79683;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=21117151; PubMed=11172064;
RA Gracey A.V., Troll J.V., Somero G.N.;
RT "Hypoxia-induced gene expression profiling in the euryoxic fish
RT Gillichthys mirabilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).
DR EMBL; AF266235; AAG13354.1; -.
DR InterPro; IPR001064; Cystallin.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
FT NON TER 1
SQ SEQUENCE 45 AA; 5435 MW; B375920482F4D2C6 CRC64;

Query Match 72.2%; Score 13; DB 13; Length 45;
Best Local Similarity 40.0%; Pred. No. 2.4e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YXXXF 5
Db 15 YASTF 19

RESULT 8
Q8KDN9 PRELIMINARY; PRT; 47 AA.
ID Q8KDN9

```

```

AC Q8KDN9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein CT1006.
GN CT1006.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103665; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey B.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Unayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vamathavan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
DR EMBL; AR012865; AAM72241.1; -.
DR TIGR; CT1006; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 47 AA; 4897 MW; 6B38DA41374AC45B8 CRC64;

Query Match 72.2%; Score 13; DB 16; Length 47;
Best Local Similarity 40.0%; Pred. No. 2.4e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 10 YTTAF 14

RESULT 9
Q8W6Q6 PRELIMINARY; PRT; 49 AA.
AC Q8W6Q6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Gp44.
GN 44.
OS Bacteriophage phiB125.
OC Viruses; dsDNA viruses.
OC Lambda-like viruses.
OX NCBI_TaxID=180504;
RN [1]
RP SEQUENCE FROM N.A.
RA Woods D.E., Jeddeloh J.A., Fritz D.F., Deshazer D.;
RT "Burkholderia thailandensis E125 Harbors a Temperate Bacteriophage
RT Specific for Burkholderia mallei.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF447491; AAL40318.1; -.
SQ SEQUENCE 49 AA; 5577 MW; 98F747E0A1D2AC53 CRC64;

Query Match 72.2%; Score 13; DB 9; Length 49;
Best Local Similarity 40.0%; Pred. No. 2.5e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 33 YAAAF 37

RESULT 10
Q8EWU8 PRELIMINARY; PRT; 56 AA.
ID Q8EWU8
AC Q8EWU8;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN MYPE1035.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300 (2002).
DR EMBL; AF004170; BAC43895.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 56 AA; 6551 MW; EB575D5D435A7A55 CRC64;

Query Match 72.2%; Score 13; DB 16; Length 56;
Best Local Similarity 40.0%; Pred. No. 2.8e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 19 YSTSP 23

RESULT 11
Q9U7H8 PRELIMINARY; PRT; 57 AA.
ID Q9U7H8
AC Q9U7H8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Erythrocyte membrane protein 1 SD106G (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99404835; PubMed=10477185;
RA Ward C.P., Clotney G.T., Dorris M., Ji D.D., Arnot D.E.;
RT recombination rearranges constrained sequences.";
RL Mol. Biochem. Parasitol. 102:167-177 (1999).
DR EMBL; AF127309; AAD52789.1; -.
FT NON_TER 1
FT NON_TER 57
SQ SEQUENCE 57 AA; 6304 MW; CD2891670148C8B6 CRC64;

Query Match 72.2%; Score 13; DB 5; Length 57;
Best Local Similarity 40.0%; Pred. No. 2.8e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 44 YASSF 48

RESULT 12
Q8CZ69 PRELIMINARY; PRT; 57 AA.
ID Q8CZ69
AC Q8CZ69;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SPB0311.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).

```

```

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McLaren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glas J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RT J. Bacteriol. 183:5709-5717(2001).
DR EMBL: AE008412; AAK99115.1; -.
DR PIR: G97910; G97910.
DR InterPro: IPR00437; Prok_lipoprot_S.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 57 AA; 6116 MW; 9FDB360477D034B5 CRC64;

Query Match 72.2%; Score 13; DB 16; Length 57;
Best Local Similarity 40.0%; Pred. No. 2.8e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 33 YSSTF 37

RESULT 13
O9PGG6 PRELIMINARY; PRT; 58 AA.
AC O9PGG6;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein Xf0336.
GN XF0336.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Ayaya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

```

```

RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AE003886; AAF83146.1; -.
DR PIR: C82818; C82818.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 58 AA; 6888 MW; 4D0944DCB5D2AA4 CRC64;

Query Match 72.2%; Score 13; DB 16; Length 58;
Best Local Similarity 40.0%; Pred. No. 2.9e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 37 YSAAF 41

RESULT 14
O8GRV3 PRELIMINARY; PRT; 59 AA.
AC O8GRV3;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Glutathione reductase (fragment).
GN RGRC2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22286393; PubMed=12399401;
RA Olsen K.M., Furugganan M.D.;
RT "Molecular Evidence on the Origin and Evolution of Glutinous Rice.";
RL Genetics 162:941-950(2002).
DR EMBL: AY136760; AAN15933.1; -.
DR EMBL: AY136761; AAN15934.1; -.
DR EMBL: AY136762; AAN15935.1; -.
DR EMBL: AY136763; AAN15936.1; -.
DR EMBL: AY136764; AAN15937.1; -.
DR EMBL: AY136765; AAN15938.1; -.
DR EMBL: AY136766; AAN15939.1; -.
DR GO: GO:0005737; C:cytoplasm; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR004099; pyr_redox_dim.
DR Pfam: PF02852; pyr_redox_dim; 1.
FT NON_TER 1 59
FT NON_TER 59
SQ SEQUENCE 59 AA; 6374 MW; B416BD7DB6B391D CRC64;

Query Match 72.2%; Score 13; DB 10; Length 59;
Best Local Similarity 40.0%; Pred. No. 2.9e+03;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5
DB 47 YTSSF 51

RESULT 15
Q9DGS1 PRELIMINARY; PRT; 60 AA.
AC Q9DGS1;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE AMVIT02.
GN AMVIT02.

```

OS Amsacta moorei entomopoxvirus (AmEPV).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;  
OC Entomopoxvirus B.  
OX NCBI\_TaxID=28321;  
RN [1]\_  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20396580; PubMed=10936094;  
RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,  
RA Moyer R.W.;  
RT "Complete Genomic Sequence of the Amsacta moorei Entomopoxvirus:  
RL Analysis and Comparison with Other Poxviruses.";  
RL Virology 274:120-139(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,  
RA Moyer R.W.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF250284; AAC02986.1; -.  
DR EMBL; AF250284; AAC02974.1; -.  
SQ SEQUENCE 60 AA; 6944 MW; FFD154D5B93AC36 CRC64;

Query Match 72.2%; Score 13; DB 12; Length 60;  
Best Local Similarity 40.0%; Pred.No. 3e+03;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YXXXF 5  
Db 14 YSSSF 18

Search completed: October 5, 2004, 16:12:35  
Job time : 27.2329 secs